Media TECHNOLOGY 传媒典藏 音频技术与录音艺术译丛

高保真音响 www.hi-fi.com.cn

ELSEVIER 爱思唯尔

音乐声学与心理声学

（第三版）

ACOUSTICS AND PSYCHOACOUSTICS

[英] David M. Howard Jamie Angus 著

陈小平 译



随书附带CD中包含78轨与本书内容相关的听音项目，通过聆听有助于更好地理解和掌握书中内容



人民邮电出版社
POSTS & TELECOM PRESS

“这本书对学音乐的学生来说是一本好书——甚至对耳科外科医生也值得一读。为了方便声学初学者阅读，书中内容通俗易懂，涉及的范围也很广泛。”

——《高保真世界》(Hi-Fi World)

本书对音乐技术、音乐录音、传统音乐和声学专业的学生来说是一本理想的学习用书，也适合从事声频、多媒体和通信系统工作的工程技术人员阅读。通过对本书富有趣味性的阅读，可获得关于乐声物理特性和乐声听觉感知的非常实用的知识。

这个新版本涉及内容广泛，包含以下主题：

- 声波原理
- 人耳听觉和心理声学
- 音色、音高和响度感觉
- 乐器声音的产生
- 不同环境中的声音（建筑声学）
- 声音的电子处理

作者简介

David M. Howard 教授在约克大学（University of York）电子系音乐技术专业任教，他的教程得到电子工程协会（Institution of Electrical engineering，IEE）的完全认可。要获得该教程的学习资格，学生必须具备音乐演奏技能以及一定的数理基础。David M. Howard的研究领域包括音乐、语声和歌声的分析与合成，人耳听觉模型的建立以及利用计算机显示屏进行歌唱训练。David还是一位热衷于管风琴演奏、男高音和合唱指挥的音乐家，他是约克大教堂的副男高音，并且担任一个12人无伴奏合唱团（The Beningbrough Singers）的指挥。

Jamie Angus 是索尔福德大学（University of Salford）计算、科学和工程学院声频技术专业的教授。他曾经在约克大学电子系任教，并和其他人一起于1986年创办了英国第一个大学水平的音乐技术教程。他的研究领域包括室内声学、语言声学、新型声扩散器设计和声频信号处理。他发明了调制式二元吸声扩散结构，并开发了可对SACD信号直接进行信号处理的技术。他休闲时喜欢以打鼓、跳舞的形式放松自己。

浏览本书的配套网页http://books.elsevier.com/0240519957，可以得到更多资讯，如：

- 检验知识掌握情况的习题
- 获取更多资料和开展进一步研究的网站链接
- 声音载片
- 计算工具（如分贝值的叠加、频率比与音分或半音数的换算公式）

这个网站也可以通过www.focalpress.com找到。





本书译自原版Acoustics and Psychoacoustics，并由Elsevier授权出版。

高保真音响 www.hi-fi.com.cn






封面设计：胡泽丽

ISBN 978-7-115-21677-9
定价：75.00 元（附光盘）

分类建议：声频/音乐技术
人民邮电出版社网址：www.ptpress.com.cn

Media
TECHNOLOGY 音频技术与录音艺术译丛
传媒典藏

音乐声学与心理声学

Acoustics and Psychoacoustics

（第三版）

[英] David M. Howard
Jamie Angus 著
陈小平 译

人 民 邮 电 出 版 社
北 京

图书在版编目（CIP）数据

音乐声学与心理声学 ：第3版 / (英) 霍华德
(Howard, D. M.)，(英) 安格斯 (Angus, J.) 著 ；陈小平
译. — 北京 ：人民邮电出版社，2010.1
（音频技术与录音艺术译丛）
ISBN 978-7-115-21677-9

Ⅰ. ①音… Ⅱ. ①霍… ②安… ③陈… Ⅲ. ①音乐声
学②心理声学 Ⅳ. ①J611.1②B845.2

中国版本图书馆CIP数据核字(2009)第194478号

音频技术与录音艺术译丛

音乐声学与心理声学（第三版）

◆ 著　[英] David M. Howard　Jamie Angus
译　陈小平
责任编辑　黄　彤
执行编辑　宁　茜

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
邮编　100061　电子函件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京隆昌伟业印刷有限公司印刷

◆ 开本：787×1092　1/16
印张：24.25
字数：475 千字　2010 年 1 月第 1 版
印数：1－3 000 册　2010 年 1 月北京第 1 次印刷

著作权合同登记号　图字：01-2009-0465 号

ISBN 978-7-115-21677-9

定价：75.00 元（附光盘）

读者服务热线：(010)67132837　印装质量热线：(010)67129223
反盗版热线：(010)67171154

内容提要

本书第 1 章介绍了声波的基本物理性质，使读者从物理学角度对声音有一定的了解；第 2 章介绍了听觉特性，即心理声学的基础知识；第 3 章到第 5 章介绍了音乐声学，主要将乐器的发声原理和音乐的听觉心理学结合起来进行讨论，在此基础上使读者对音乐有更深层次的理解，为从事音乐技术工作打下基础；第 6 章主要介绍了声波在封闭空间即房间中的传播特性，使读者了解室内声学特性和音质之间的密切关系，达到对室内音质进行控制的目的；第 7 章介绍了声音信号电子加工和处理的几种基本方法。

本书还附带有一张 CD，录制了 78 条听音项目，目的是使读者通过听音更好地理解书中的相关内容。

本书适合从事音响工程、音乐技术、音乐录音、音乐制作的人员阅读，通过阅读可以提高专业理论水平，对实际工作将有较大的促进作用，也适合从事相关教学和学习的院校师生作为一本很好的参考书。

音频技术与录音艺术译丛编委会

丛书编委会主任简介



李伟，生于1952年，沈阳人。20岁在沈阳音乐学院舞台美术系灯光音响专业学习，毕业后留校任教并兼做扩声和录音工作，之后赴德国柏林艺术大学（UdK）音乐学院录音专业攻读硕士学位，师从录音大师J. N. 马蒂斯教授。学成回国后调入中国传媒大学（原北京广播学院）任教，出版专著《立体声拾音技术》、主编《现代录音技术丛书》。现任中国传媒大学影视艺术学院副院长、录音系系主任、硕士研究生导师，德国录音师协会会员，中国电影电视技术学会声音专业委员会专家组成员，国家广电总局“电视节目质量奖”（金帆奖）评委，中国电影电视技术学会“声音制作优秀作品奖”评委，亚洲录音艺术与科学（广州）文化节“音乐录音作品”评委。

总　　序

翻译一套现代录音技术丛书是我们多年的夙愿。

随着现代科技的不断进步和现代媒体传播形式的不断演变，现代录音技术的发展也是十分迅猛的。我国在声音设计与制作领域的理论研究与实践近些年来取得的成就是有目共睹的，尤其是 2008 年北京奥运会成功举办，高清电视转播和环绕声声音制作使电视声音制作水平提高到新的阶段。但是，与欧美发达国家相比较，我国在该领域还存在一定的差距。中国传媒大学影视艺术学院录音系，作为国内从事声音方面理论研究和教学的团队，一直关注和跟踪国外该领域的研究动态和实践成果，并将国外该领域的许多专著的思想和方法注入到我们的教学中。"它山之石，可以攻玉"，如果将国外最新的录音技术专著翻译出版，无疑是一件很有意义的事情，于是，我们萌生了翻译一套录音技术丛书的计划。

2007 年夏天，有幸结识了人民邮电出版社《高保真音响》杂志社的黄彤主编和宁茜编辑。他们十分支持我们的计划，并提供了 Focal Press 的最新书目。对于这套丛书的设计思想、读者定位等我们也是一拍即合，于是，我们录音系的全体老师带领部分研究生开始了紧张的丛书翻译工作。

今天，"音频技术与录音艺术译丛"的第一批译著与读者见面了，其他译著将分批陆续出版。这套丛书包括广播、电影、电视、唱片等领域的录音技术基础理论、系统集成、声音设计、拾音方法、制作技巧等方面，内容十分丰富，甚至有些译著涉及的领域是国内目前出版物的空白。但愿这套丛书能够为广大从事声音设计和制作的专业人士、业余爱好者和本专业的学生提供帮助，也希望广大读者对本丛书的设计、翻译等诸方

面提出宝贵意见。

感谢人民邮电出版社副社长蒋伟先生，他曾亲自到我系就本套书的出版事宜进行指导。感谢黄彤主编和宁茜编辑，正是他们的大力支持和高效工作使得这套丛书成功出版。感谢录音系的团队，是我们共同营造的宽松的学术氛围、严谨的治学精神和兄弟姐妹般的情谊使这套书能够顺利地翻译完成。

中国传媒大学影视艺术学院副院长、录音系系主任

李 伟

2008 年初冬于北京

译者简介

陈小平，1963 年出生，1983 年毕业于北京邮电学院电信工程系，并获得工学学士学位，1988 年获得北京广播学院通信与信息系统专业广播声学方向工学硕士学位。1999 年 7 月至 2000 年 7 月，在丹麦 Aalborg 大学声学系任访问学者。现在中国传媒大学录音系任教，承担课程有《声学基础》、《电声学与室内声学》等，研究领域为通信声学（Communication Acoustics），主要研究方向为心理声学与听觉传输技术（Phsychoacoustics and Binaural Technology）。



前　　言

前两个版本在全世界拥有大量的读者，说明此书的出版是一个巨大的成功。一直以来我们不断收到读者来信，对本书有关内容提出了一些小的意见和建议（我们十分欢迎这样的交流！）。在第三版的修订中，我们充分考虑了这些意见和建议，对本书内容进行了完善。我们对这些读者表示衷心的感谢！

第三版中补充了一些内容，其中最值得一提的是增加了一张 CD，为利用本书进行教学工作提供了有用的录音资料。在制作这张 CD 时，我们参考了由美国音响工程协会提供的非常优秀的专门用于听音演示的 CD（Philips CD No. 1126-061），这是由 Houtsma，A.J.M.，T.D. 和 Wagenaars，W.M. 制作的。尽管之前我们在使用本书作为教材时选用的是这张 CD，并且推荐他人使用该 CD 作为本书教学的补充，但是我们也意识到仅仅简单地重复使用该 CD 的内容并不合适。因此，在第三版所附的 CD 中，我们增加了一些听音材料和单独录制了一系列的单音符、在消声室录制了由特别选择的不同乐器演奏的短的音乐片断（只有直达声而不存在由房间墙壁产生的反射声），这些可以作为研究和了解不同乐器直达声的声学特性和听觉特性与其声压—时间波形和振幅—频率的频谱之间关系的源材料。附录 5 给出了该 CD 中所有听音项目一览表，指出了每个听音项目所支持的章节以及对本听音项目的一些说明。

在第三版中还对大多数章节增加了一定的内容，其一是使之跟上技术发展的潮流，其二是试图对一些问题给出更好、更清楚的说明。这些内容有些涉及非常重要的研究课题，有些则是横跨科学和艺术两个领域的，因此需要从科学和艺术两个角度进行解释。这种多学科的分析方法

对全面正确地理解音乐声学和听觉心理学是必不可少的，我们希望第三版在满足感兴趣读者的学习要求上能向前迈进一大步。

我们不断收到读者来信提出意见和建议，并且希望继续收到读者直接寄给我们或寄给出版社的来信。我们一直在这个领域从事研究工作，并从读者的反馈中得到了一些很深的启示。我们相信，对音乐和音乐技术中的声学和心理声学了解得越多，对这一人类交流的最基本形式就有更深刻的了解。

David M. Howard（York），Jamie Angus（Salford）

2005年12月

目　　录

第 1 章　声音

声音是大自然中存在的一种现象，没有人会否认它的存在。我们的周围充满了噪声，甚至在我们出生之前就已经置身于噪声的包围之中。什么是声音？声音如何传播？如何定量表示声音？本章的主要目的是介绍声音的基本要素、声音的传播方式以及其他有关问题。在此基础上，我们才能进一步了解声音的基本性质及其在不同声学环境中的传播特性，并了解乐器工作的声学原理以及声音与人耳听觉之间的相互作用关系。

1.1　压力波和声音的传播

从物理学上说，声音实际上是媒质的机械扰动。这些媒质可以是空气、固态物质、液态物质或其他气体。然而，这种简单的描述不足以说明媒质的扰动到底是怎样传播的，以及为了能够传播，媒质还需要具备其他哪些性质。因此，关于什么是声音需要一个更准确的描述，在此基础上才能够分析并预测声音在不同声学环境中的传播和存在方式。

1.1.1 声波的性质

图 1.1 所示为声音通过某种媒质传播的简单机械模型。它可以表示空气媒质简化的一维模型，是由一系列代表质量的高尔夫球通过弹簧连接在一起，称之为高尔夫球弹簧模型。高尔夫球代表小体积媒质分子（称为媒质质点）的实际质量，弹簧则代表媒质分子之间存在的相互作用力。当一端的高尔夫球受力向另一端推移时，相连的弹簧会被压缩并推动相邻的高尔夫球，这个高尔夫球又会压缩下一个弹簧，依此类推。由于高尔夫球具有一定质量，从弹簧施加作用力到高尔夫球产生移动存在一个延迟时间，因此，扰动从第一个球传递到另一端需要一定时间。如果始端的球移动回到原来的位置，则上述的运动过程也会发生，所不同的是，高尔夫球是被拉回而不是被推移，弹簧是被拉伸而不是被压缩。无论哪一种情况，当运动停止时，高尔夫球之间的间距会恢复到原来静止状态的间距。



图 1.1 声音传播媒质的高尔夫球弹簧模型

高尔夫球被推到一起的区域称为压缩区，而高尔夫球被分离开的区域称为稀疏区，高尔夫球本身就是振动传播的媒质。在真实媒质如空气中，由于空气媒质质点总是要回到平衡位置，扰动往往是由压缩状态到稀疏状态再到压缩状态的不断反复过程，如图 1.2 所示。高尔夫球沿着媒质扰动传播的方向前后移动，这种波动传播方式称为纵波。由于声波是通过媒质（大多数情况下是空气媒质）的一系列压缩和稀疏过程进行传播的，因此属于纵波。

另一种媒质扰动方式是沿着与高尔夫球弹簧系统垂直的方向进行的，使高尔夫球不是被前后推拉，而是上下移动。由于弹簧对高尔夫球的作用，这种横向扰动也会得到传播，这种波动的传播方式称为横波。乐器中的振动往往属于这一类，例如琴弦或鼓膜的振动。

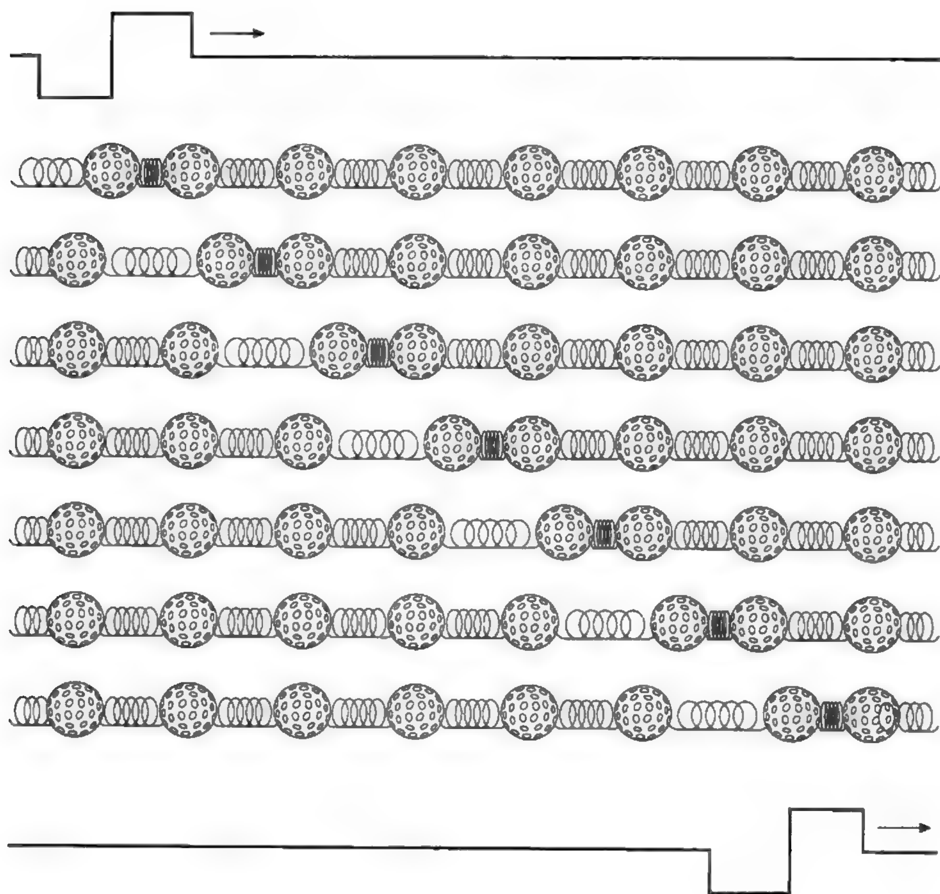


图 1.2　声脉冲在媒质中传播的高尔夫球弹簧模型

1.1.2　声波传播速度

无论是横波还是纵波，扰动沿着高尔夫球弹簧连成的“弦”向前移动的速度取决于以下两个方面。

（1）高尔夫球的质量：由于高尔夫球质量越大，球开始或停止移动所需的时间越长，因此质量影响扰动传播的速度。实际上，媒质的密度决定了高尔夫球的有效质量。媒质密度越大则球的有效质量越大，那么扰动传播得越慢。

（2）弹簧的强度：连接高尔夫球的弹簧强度也会影响扰动的传播速度。因为弹簧的强度越大，则对相连的球的作用力越大，使之移动的速度就越快。实际上媒质的弹性强度取决于媒质的弹性模量，即所谓的杨氏模量。媒质的弹性模量越高，意味着弹簧的强度越大，扰动的传播速度就越快。[注 1]

对固体媒质中的纵波而言，传播速度仅由媒质的密度和杨氏模量决定，可简单地用下列公式计算：

$$v=\sqrt{\frac{E}{\rho}} \tag{1.1}$$

其中，v 为扰动每秒传播的距离（m/s），ρ 为媒质密度（kg/m^3），E 为媒质的杨氏模量（N/m^2）[注 2]。

虽然固体媒质的密度与传播方向无关，但是杨氏模量却不一定如此。例如，黄铜是各向均匀的介质，因此其杨氏模量与传播方向无关；而木材则不然，其杨氏模量大小取决于是沿木材纹理方向测量还是沿横穿纹理方向测量。因此，黄铜中扰动的传播速度与方向无关，而木材中传播速度则与扰动是沿着木材纹理方向还是沿着横穿纹理方向传播有关。下面举例说明。[注 3][注 4][注 5]

例 1.1　试计算声波在钢材和榉木中的传播速度。

钢材的密度是 7 800kg/m^3，其杨氏模量为 2.1×10^{11}N/m^2，则声波在钢材中的传播速度为：

$$v_{st}=\sqrt{\frac{2.1\times10^{11}}{7\,800}}=5\,189\ \text{m/s}$$

榉木的密度是 680kg/m^3，其杨氏模量沿纹理方向是 14×10^9N/m^2，沿横穿纹理方向是 0.88×10^9N/m^2。因此，声波沿不同方向传播速度不同，分别是：

$$v_{be1}=\sqrt{\frac{14\times10^{9}}{680}}=4\,537\ \text{m/s}$$

$$v_{be2}=\sqrt{\frac{0.88\times10^{9}}{680}}=1\,138\ \text{m/s}$$

可见，沿榉木纹理方向的声速约为沿横穿纹理方向的 4 倍。

木材这种传播速度随传播方向变化的特性往往会影响木制乐器的声学性能。由于扬声器箱体大多数是木制的，因此在设计时也要考虑到这种因素的影响。通常扬声器制造商选用加工过的木材如胶合板、MDF（中密度板）等制作箱体，这些材料的杨氏模量不随方向变化，因此可以不考虑这个因素对箱体设计的影响。

[注 1] 杨氏模量是对材料弹性的度量，较高的杨氏模量意味着需要更大的力使材料压缩变形。其度量单位是牛顿每平方米（N/m^2）。

[注 2] 牛顿（N）是力的度量单位。

[注 3] 密度是单位体积的质量，单位是千克每立方米（kg/m^3）。

[注 4] $\sqrt{\ }$ 是平方根的运算符号，它表示对里面的数求平方根。

[注 5] 1/s 意指每秒。

1.1.3　空气中的声速

到目前为止我们仅讨论了固体中的声速。然而，最常见的声音是在空气中传播的，音乐就是通过空气传播到人耳的。虽然空气传播声音的机理与固体完全相同，但由于空气不存在杨氏模量，因此式（1.1）不能直接应用于空气中声速的计算。但空气是有弹性的，当你使用自行车打气筒时你能感觉到这一点，因此需要找到空气的某个参数取代杨氏模量。在绝热情况下，气态方程为：

$$PV^{\gamma}=\text{常数} \tag{1.2}$$

其中，P 为气体压强 [注 1]（N/m^2），V 为气体体积（m^3），γ 为由气体性质决定的常数（空气中为 1.4）。

通常媒质的扰动变化很快，以至没有足够的时间进行稠密位置与稀疏位置之间的热交换，因此上述气体绝热方程是适用的。式（1.2）给出了气体的压强和体积之间的关系，由此式可导出空气弹簧的强度，即空气的等效杨氏模量为：

$$E_{gas}=\gamma P \tag{1.3}$$

气体的密度由下式给出：

$$\rho_{gas}=\frac{m}{V}=\frac{PM}{RT} \tag{1.4}$$

其中，m 为气体的质量（kg），M 为气体的摩尔质量（kg/mol）[注 2]，R 为气体常数（8.31J/(K·mol)），T 为气体的绝对温度（K）。

由式（1.3）和式（1.4）可得空气中的声速计算公式为：

$$v_{gas}=\sqrt{\frac{E_{gas}}{\rho_{gas}}}=\sqrt{\frac{\gamma P}{\left(\frac{PM}{RT}\right)}}=\sqrt{\frac{\gamma RT}{M}} \tag{1.5}$$

式（1.5）的重要之处在于它说明了气体媒质中的声速与压强无关，而与绝对温度和气体的摩尔质量有很大关系。因此我们可以认为，轻质气体如氦气中的声速要比摩尔质量较大的气体如二氧化碳中的声速大，空气中的声速应介于二者之间。空气中声速的计算用以下例子说明。

例 1.2　试计算 0℃和 20℃时空气中的声速。

空气中含 21% 的氧（O_2）、78% 的氮（N_2）、1% 的氩（Ar）和微量的其他气体，因此空气的摩尔质量可计算如下：

$$M=21\%\times16\times2+78\%\times14\times2+1\%\times18=2.87\times10^{-2}\text{kg/mol}$$

由于

$$\gamma=1.4$$
$$R=8.31\text{J/(K·mol)}$$

所以，声速为：

$$v=\sqrt{\frac{1.4\times8.31}{2.87\times10^{-2}}T}=20.1\sqrt{T}$$

因此，空气中的声速仅与绝对温度的平方根成正比，而绝对温度值可由摄氏温度加 273 得到，因此 0℃和 20℃时空气中的声速分别为：

$$v_{0℃}=20.1\sqrt{(273+0)}=332\text{m/s}$$

$$v_{20℃}=20.1\sqrt{(273+20)}=344\text{m/s}$$

声速随温度的上升而增大可以从两个方面来解释。第一，由计算理想气体密度的式（1.4）可知，当温度上升时，体积会随之增大，只要压强保持不变，则气体密度减小。第二，如果压强确实发生了变化，它对密度的影响会被空气中等效杨氏模量的提高所抵消，参看式（1.3）。事实上，除了温度以外，对气体中声速产生影响的另一个主要因素是气体的摩尔质量。显然，不同的气体具有不同的摩尔质量，例如，氦气的摩尔

质量就与空气不同。因此，当空气中有水蒸气存在时，摩尔质量会发生变化，因为水分子取代了部分空气分子。由于水分子较轻，因此导致声速比在干燥环境下的声速略有提高。

虽然空气中的声速与绝对温度的平方根成正比，我们可以导出声速与摄氏温度之间大致的线性关系为：

$$v=331.3+0.6t \text{m/s} \quad (1.6)$$

其中，t 为空气的摄氏温度（℃）。

由此可以看出，温度每升高 1℃，声速会增加 0.6m/s。因此温度对声波的传播方式有较大影响。

表 1.1 给出了不同材料的密度、杨氏模量和纵波的传播速度。

表 1.1　常用材料的杨氏模量、密度和声速

材料	杨氏模量（N/m^2）	密度（kg/m^3）	声速（m/s）
钢	2.10×10^{11}	7 800	5 189
铝	6.90×10^{10}	2 720	5 037
铅	1.70×10^{10}	11 400	1 221
玻璃	6.00×10^{10}	2 400	5 000
混凝土	3.00×10^{10}	2 400	3 536
水	2.30×10^{9}	1 000	1 517
空气（20℃）	1.43×10^{5}	1.21	344
榉木（沿纹理）	1.40×10^{10}	680	4 537
榉木（横穿纹理）	8.80×10^{8}	680	1 138

[注 1] 压强表示气体施加于某个单位面积表面的作用力，力的单位是牛顿（N）。作用力产生于气体分子对表面的撞击。压强单位是牛顿每平方米（N/m^2）。

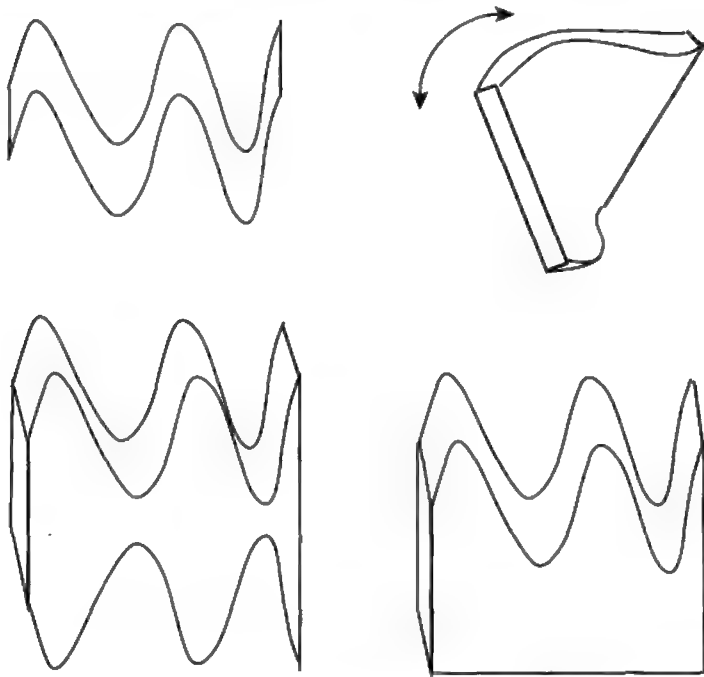
[注 2] 气体分子的质量约等于分子中包含的所有质子和中子质量的总和，单位是克（g）。一个摩尔（mol）相当于包含 6.022×10^{23} 个分子。

1.1.4　横波传播速度

横波的传播速度受其他因素影响。例如，弹簧的静态弹性力对高尔夫球弹簧模型中高尔夫球的加速度有很大影响。当弹性力较小时，高尔夫球回到平衡位置的恢复力也较小，则波动传播的速度比弹性力较大时慢。同时，对于三维物体，可能存在不同形式的横波。例如，振动方向

可能不同，而且由于对面的振动状态同相或者反相，会产生不同的波动形式，如横波、扭曲波以及其他形式的波，如图 1.3 所示。由于这些不同的波动形式具有不同的弹性系数，而且不同程度地受外界因素如物体形状等影响，因此，除一根细弦这样简单形状的物体外，任何其他形状物体的横波波速问题都变得极其复杂。当需要分析打击乐器的声学特性时，就要遇到这些复杂问题。

图 1.3　一些不同形式的横波



对于横波，波速的计算会比较复杂，因为正如前面提到的，除了截面积无限小的细弦外，任何媒质中横波波速都会受到传输媒质的形状和横波形式的影响。然而，许多乐器的振动属于细弦的横振动。细弦中横波的传播速度可用下式计算：

$$v_{tr}=\sqrt{\frac{T}{\mu}} \tag{1.7}$$

其中，μ 为弦的单位长度质量（kg/m），T 为弦的张力（N）。

虽然公式推导时假设弦的截面积为无限小，但是上述公式适用于大多数实际应用中的弦。需要注意的是，它仅适用于理想的横振动，而不

适用于扭曲等其他形式的横振动。所幸的是细弦的横振动是以理想横振动为主的，其产生传播速度计算误差的主要原因是：由于实际材料中存在一定的硬度，使得传播速度随频率的升高略有提升。这种现象不会影响打击弦乐器如钢琴的音色，但对较粗弦的音色影响会变大。因此，对于大多数实际情况而言，式（1.7）是适用的。下面以钢弦为例计算横波波速。

例 1.3 试计算直径为 1.8mm、张力为 627N 的钢弦（可能是一根吉他弦）的横振动传播速度。

钢弦单位长度质量为：

$$\mu_{steel}=\rho_{steel}\left(\pi r^2\right)=7\,800\times 3.14\times\left(\frac{0.8\times 10^{-3}}{2}\right)^2=3.92\times 10^{-3}\text{kg/m}$$

因此，横波的波速为：

$$v_{tr}=\sqrt{\frac{627}{3.92\times 10^{-3}}}=400\text{m/s}$$

这一数值远小于同一材料中纵波的传播速度。一般来说，同一种材料的横波传播速度比纵波传播速度小。

1.1.5 声波的波长和频率

目前为止，我们仅考虑了一次性的、单一的位移扰动在高尔夫球弹簧模型中的传播情况，并且看到扰动以恒定的速度传播，且传播速度仅与媒质的性质有关。因此，任何其他形式的扰动，例如周期性的扰动，也将以恒定的速度传播。图 1.4 所示为受恒速转动的轮子驱动的高尔夫球弹簧模型。模型的一端用大头针固定在轮子上，这将产生一个随时间按正弦规律变化的压强，正弦函数的相位角就是轮子转动的角度。这种激励称为正弦激励，由此产生正弦波。正弦波代表最简单的周期性激励信号，在本书的讨论中十分重要。正弦波有三个重要参数，即振幅、转速或频率和初始状态或初相位。频率的单位是每秒周数，现在常用赫兹（Hz）表示，它反映了波形的基本特点。这种激励会沿着模型产生向前推进的正弦波扰动，使压缩和稀疏按周期性规律变化。由于正弦波的传播速度恒定不变，因此，在任何相邻的两个压缩或稀疏状态之间存在一定的距离，

如图 1.5 所示。而且由于传播速度恒定，两个相邻重复状态之间的距离与正弦波变化的速率即频率成反比，这一距离是非常重要的声学参数，称为波长（一般用 λ 表示）。波长、频率和传播速度三者密切相关，知道其中两个量可以计算出第三个量，计算公式如下：

$$v=f\lambda \quad (1.8)$$

其中，v 为媒质中的声速（m/s），f 为声波的频率（Hz，1Hz = 1 周每秒），λ 为媒质中的波长（m）。

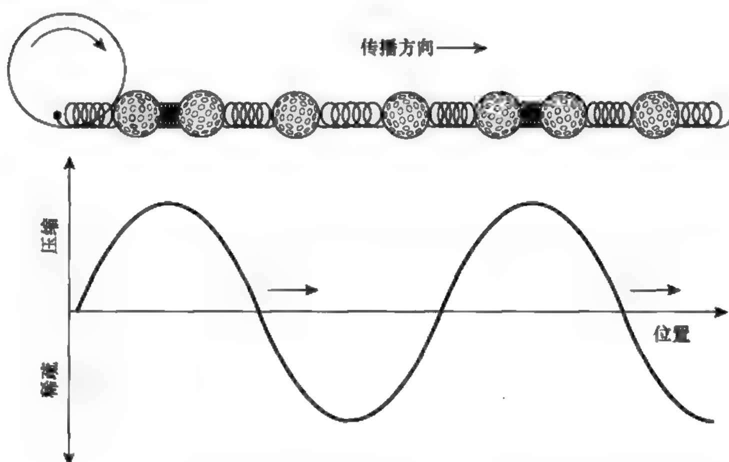


图 1.4　正弦波在媒质中传播的高尔夫球弹簧模型

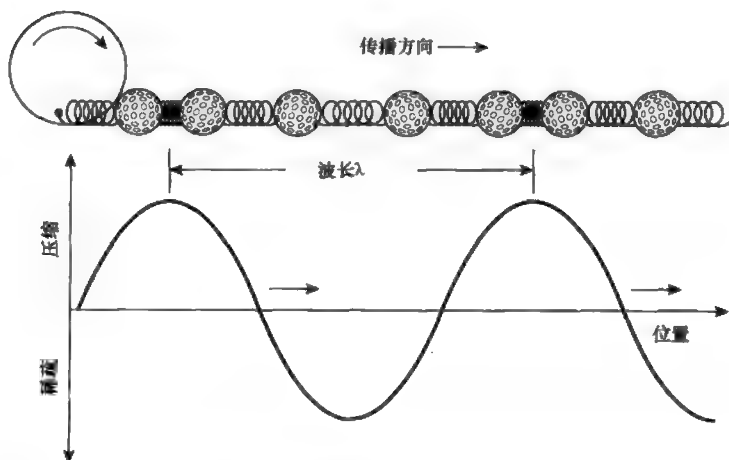


图 1.5　正弦波的波长

上式可用于已知波长计算频率，或已知频率计算波长，甚至用于已知频率和波长计算媒质中的声速，同时适用于纵波和横波。

例 1.4　试计算 20℃空气中，频率为 20Hz 和 20kHz 的声波波长。

20℃空气中的声速是 344m/s（见例 1.2），因此，两种频率下的波长可按下式计算：

$$\lambda = \frac{v}{f}$$

因此得到：

$$20\text{Hz 时，}\lambda = \frac{344}{20} = 17.2\ \text{m}$$

$$20\text{kHz 时，}\lambda = \frac{344}{20 \times 10^3} = 1.72\ \text{cm}$$

这两个频率分别对应声频的下限和上限频率，由此可见声频所包含的波长范围非常之大！

例 1.5　试计算 20℃空气中波长为 34cm 的声波频率。

频率可由下式计算：

$$f = \frac{v}{\lambda} = \frac{344}{0.34} = 1\ 012\ \text{Hz}$$

在声学中，波长通常作为衡量长度的尺度，而不使用米、英尺或弗隆（译者注：1 弗隆相当于 201 米或 1/8 英里）。因为声波在实际空间如房间中或遇到障碍物时的传播特性往往与波长有关。

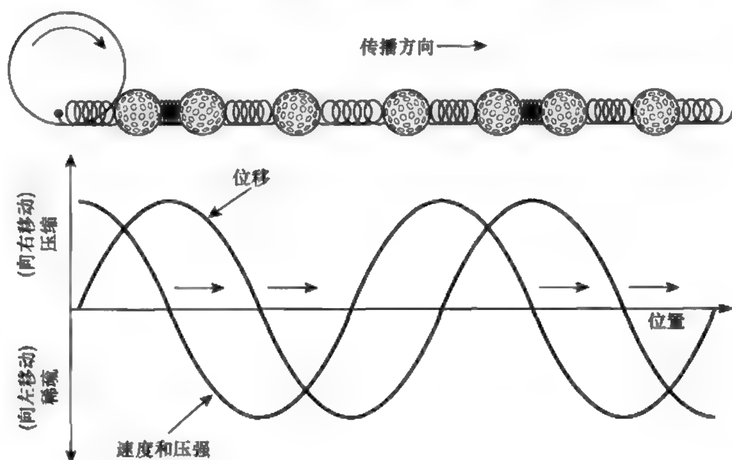
1.1.6　声波中声压、振速和阻抗的关系

声波中需要考虑的另一个问题是传播声波的媒质质点的运动情况。声波可看成一系列沿着媒质传播的压缩和稀疏变化，这种效应是由媒质质点位移和加速度产生的。使质点产生位移的作用力也使声波具有一定的声压。

为了使媒质产生压缩和稀疏状态，媒质质点需要通过运动使质点之间更加紧密或进一步分离。运动则存在速度，因此，必然存在一个与声波的位移量相关的速度量。这一现象可以从前述的高尔夫球模型看出：为了使高尔夫球更加紧密而产生媒质的压缩状态，高尔夫球必须以某一

速度移动而彼此靠近。当压缩最大时，速度为零，质点处于静止状态，然后质点开始向相反方向移动以产生稀疏状态，同样当稀疏达到最小时质点的速度又变为零。由于质点运动存在惯性，质点的速度不会突然从一个方向变到另一个方向，而是连续地以一定的加速度从零变到一定的速度值，然后再回到零。速度在压缩状态和稀疏状态之间达到最大值。如果质点位移按正弦规律变化，则对应的质点速度按余弦规律变化。图1.6所示为按正弦规律传播的高尔夫球弹簧模型及其速度和位移曲线。使质点产生加速度的力产生了声波的压强变量，因此压强与声波的速度变量有关，与速度是同相位的，即如果速度按余弦规律变化，则压强也是按余弦规律变化。声波的压强和速度是以相同的速度沿着媒质传播的。

图 1.6　某媒质中正弦波的压强、速度和位移曲线



压强是标量，没有方向性。我们在谈到压强时通常是指某一点的压强，而不是指某一特定方向的压强。相反，速度具有方向性，物体总是沿着某一方向运动。速度的方向决定了声波的传播方向。

声波的质点速度和压强大小是通过传播媒质的密度和弹性彼此相联系的。如果媒质具有较小的密度和较弱的弹性，在压强振幅相同的情况下，其速度振幅要比在较大密度和较强弹性的媒质中大。对于距声源和所有边界较远的声波，质点速度和压强的关系可用下式表示：

$$\frac{\text{声压振幅}}{\text{速度振幅}} = \text{常数} = Z_s = \frac{p}{u} \qquad (1.9)$$

其中，p 为声压振幅，u 为速度振幅，这一常数称为声阻抗率，用 Z_s 表示，

与电路中的电阻（或阻抗）相类比。

声压振幅是媒质弹性（杨氏模量）的函数，速度是媒质密度的函数。用杨氏模量和密度计算声阻抗率的公式为：

$$Z_s = \sqrt{\rho E}$$

而声波的传播速度 c 也与媒质的杨氏模量和密度有关，因此，在 20℃的空气中，上式通常表示为：

$$Z_s = \sqrt{\rho E} = \sqrt{\rho^2 \left(\frac{E}{\rho}\right)} = \rho c = 1.21 \times 344 = 416\text{kg/(m}^2\text{·s)[注 1]} \qquad (1.9a)$$

因此，在自由空间（无边界）的情况下，声阻抗率也仅与传播媒质的特性有关。

但是，当声波在管中传播，且管的截面尺寸小于波长时，式（1.9）要用管的截面积进行修正，得到声阻抗（译者注：声阻抗等于声阻抗率除以面积，一般用 Z_a 表示）计算公式为：

$$Z_{\text{atube}} = \frac{\rho c}{S_{\text{tube}}} \qquad (1.9b)$$

其中，S_{tube} 为管的截面积。

这意味着有界波的声阻抗由边界的截面积决定，并随着截面积的变化而变化。在后面的章节我们将会看到，阻抗的变化会使声波产生反射。这个现象的认识对理解第 4 章中介绍的各种乐器的设计和工作原理是非常重要的。

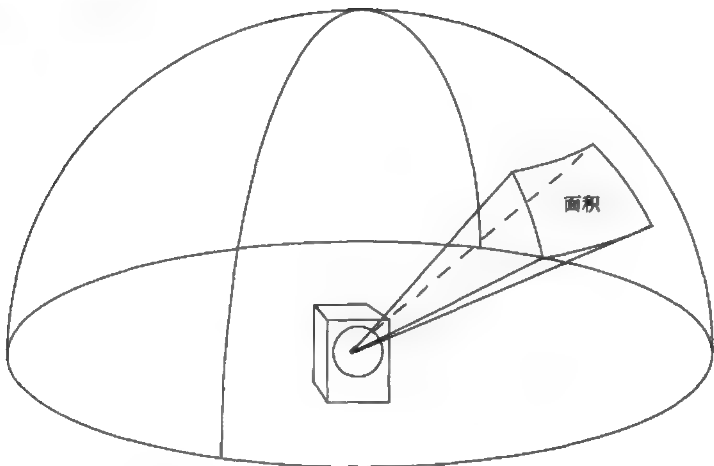
[注 1] $1/m^2$ 表示每平方米。

1.2 声强、声功率和声压级

声波的能量用于度量所存在声波的量的大小。而一般情况下，我们对声能量传输的速率比对所传输的总能量更感兴趣。因此，我们对单位时间内传输的声能量即每秒焦耳数（W）感兴趣。声波也是三维的，它占据一定的空间，因此，在描述声能量速率时常引入面积的概念，即用单位面积的瓦特数来描述声能量速率。由此得到声强这个物理量（译者注：Sound Intensity，一般用 I 表示），用来度量声波在某个方向传播的功

率密度，如图 1.7 所示。

图 1.7 声强



1.2.1 声强级

声强表示通过单位面积的声能量流。换句话说，它表示穿过单位面积的来自声源的声功率，这意味着声强可由声功率除以声源辐射的总面积得到。正如前面提到的，声强具有方向性，其方向垂直于能量流穿过的面积，如图 1.7 所示。实际声源的声强可能在 1 到大于 10^{12} 的范围内。由于这个原因以及人耳对声音响度的感觉特性，声强通常用其对数值即声强级来表示。该度量方式基于声波的实际功率密度和 1 微瓦每平方米（10^{-12}W/m^2）参考功率密度的比值。声强级（Sound Intensity Level，缩写为 SIL）定义为：

$$SIL = 10\lg\left(\frac{I}{I_{ref}}\right) \quad (1.10)$$

其中，I 为实际声功率密度（W/m^2），I_{ref} 为参考声功率密度（10^{-12}W/m^2）。

为了使声强级的一个整数变化对应于人耳能够感觉到的最小音量变化，采用 10 作为乘系数。声强比变化 10 倍称为一个贝尔（Bel，音量比率单位），则由式（1.10）得到声强级将变化 10；当声强变化 $\sqrt[10]{10}$ 倍时，声强级变化为 1。因此，声强级的单位取为分贝（decibel 或 dB）。$\sqrt[10]{10}$ 约等于 1.26。

例 1.6 某直径为 25cm 的扬声器辐射声功率为 20mW，试问其在扬声器处的声强级是多少？

由于声强是单位面积的声功率，因此首先求出扬声器的总辐射面积为：

$$A_{speaker}=\pi r^2=\pi\left(\frac{0.25}{2}\right)^2=0.049\text{m}^2$$

那么，声强为：

$$I=\left(\frac{W}{A_{speaker}}\right)=\left(\frac{20\times10^{-3}}{0.049}\right)=0.41\text{W/m}^2$$

将声强计算结果代入式（1.10）得声强级为：

$$SIL=10\lg\left(\frac{I}{I_{ref}}\right)=10\lg\left(\frac{0.41}{10^{-12}}\right)=116\text{dB}$$

1.2.2 声功率级

声功率级（Sound Power Level）是对声源向各个方向辐射出的总功率的度量，通常用缩写 SWL 表示，有时也用 PWL 表示。声功率级也是用比率的对数值度量的，单位为分贝（dB），即用实际声功率与 1 微瓦（10^{-12}W）的参考声功率的比值由下式计算：

$$SWL=10\lg\left(\frac{w}{w_{ref}}\right) \tag{1.11}$$

其中，w 为实际声功率（W），w_{ref} 为参考声功率（10^{-12}W）。

声功率级主要用于评估某声源辐射的总声功率，例如，用于比较不同声源产生的干扰噪声大小。从第 6 章我们将看到，声功率本身具有与声学环境无关的优点。同时请注意，声功率不具有方向性，这一点与声强有所不同。

例 1.7 某声源辐射总声功率为 1W，试计算其声功率级。

将声功率代入式（1.11）得：

$$SWL=10\lg\left(\frac{w}{w_{ref}}\right)=10\lg\left(\frac{1}{10^{-12}}\right)=10\lg\left(10^{12}\right)=120\text{dB}$$

1W 的声功率如果全部被接收到，所呈现的声压级会非常高，声音的响度会非常大。然而，在大多数情况下，听音者只能接收到小部分的声源声功率。

1.2.3 声压级

声强是度量和描述声波在某一点幅度大小的一种方式，虽然在理论上具有一定的价值，并且能够度量声波的强弱，但是，它不是用于描述声波大小的常用量。其他用于描述声波的量还有声压振幅和与之相关的媒质质点振速振幅。由于人耳对声压非常敏感（这方面内容将在第 2 章进行介绍），并且声压易于测量，通常使用声压作为描述声波大小的物理量。声压定义为声波在某一点产生的逾量瞬时压强的均方根值（即有效值）。实际声源产生的声压可以在小于 20μPa 到大于 20Pa 之间 [注 1]，其中 1Pa 等于 1 牛顿每平方米（$1N/m^2$）。这两个声压值大体上分别对应于人耳听觉在 1kHz 时的听阈（20μPa）和痛阈（20Pa）。因此，实际声源声压的变化范围可达 10^6。正由于此，而且由于人耳对声音的感觉特性，声压通常也用其对数值表示，这种度量方法是建立在实际声压与 20μPa（大约为 1kHz 人耳听阈）比值的基础上。声压级（Sound Pressure Level，缩写为 SPL）定义为：

$$SPL = 20\lg\left(\frac{p}{p_{ref}}\right) \tag{1.12}$$

其中，p 为实际测得的声压（Pa），p_{ref} 为参考声压（20μPa）。

用 20 作为乘系数主要有两个目的。其一是使声压级的一个整数变化（即变化 1 时）对应于人耳所能感觉到的声压的最小变化量；其二是使声压级与声强级的度量具有一致性，这一点说明如下。

声强可由速度和声压相乘得到：

$$I = up$$

其中，p 为声压幅度，u 为速度幅度。

然而，声压幅度和速度幅度是通过声阻抗率联系起来的（见式（1.9）），所以，声强可由声压和声阻抗率进行计算：

$$I = up = \left(\frac{p}{Z_s}\right)p = \frac{p^2}{Z_s}$$

因此，声强级可由声压幅度和声阻抗率按下式计算：

$$SIL = 10\lg\left(\frac{I}{I_{ref}}\right) = 10\lg\left(\frac{\frac{p^2}{Z_s}}{I_{ref}}\right) = 10\lg\left(\frac{p^2}{Z_s I_{ref}}\right)$$

上式说明声强与声压的平方成正比，就像电功率与电压的平方成正比一样。在对数中，平方运算可以转换为对数值乘以 2，由此得到：

$$SIL = 20\lg\left(\frac{p}{\sqrt{Z_s I_{ref}}}\right)$$

上式与式（1.12）很相似，只是参考声压以不同的形式表示。事实上，当参考声压按下式计算时：

$$p_{ref} = \sqrt{Z_s I_{ref}} = \sqrt{416 \times 10^{-12}} = 20.4 \times 10^{-6}\ (\text{Pa})$$

则声强级的计算公式与式（1.12）基本相同。上述参考声压与 20μPa 十分接近，足以认为声压级和声强级这两个度量值大体相等，即在单一声源的条件下，且与声源和任一边界相距足够远时，*SIL* 约等于 *SPL*。声强级和声压级可以是等效的，是因为声压级是针对某一点计算的，而声强级也是声源在某一点产生的声功率密度。然而，不同之处是，声强级是某个声源在测量点产生的功率密度，而声压级可以是多列声波在测量点产生的声压总和。当只有一列声波存在时，或者说在不存在任何反射声的情况下，声压级和声强级可以近似认为是相等的，即 *SPL* 等于 *SIL*。这种情况可以指测量点远离所有反射面的情况。当处在某个房间里时，则存在多个反射波，其声阻抗会发生变化，声压级和声强级则不相等，但是声压级和声强级的变化保持相等，因为当某一点的声强增大时，该点的声压会随之增大相同的比例，只要声波的数量和传播方式没有因为某种因素而改变。因此，10dB 的声强级变化将产生 10dB 的声压级变化。

这两个描述和度量声波大小的量很容易混淆，必须弄清在什么情况下使用哪一个量。当问题与声音大小有关时，通常使用声压级来度量，因为声压级较容易测量，并且与听觉听到的声音大小关系最为密切。

下面举例计算不同声压的声压级。

例 1.8 试分别计算声压有效值为 1Pa、2Pa 和 2μPa 的声波的声压级。

将已知数代入式（1.12）得：

$$SPL_{1Pa}=20\lg\left(\frac{p}{p_{ref}}\right)=20\lg\left(\frac{1}{20\times10^{-6}}\right)=20\lg\left(5\times10^{4}\right)=94\text{dB}$$

1Pa 通常用于传声器灵敏度指标的测量。从上述计算结果可知，1Pa 对应于很大的声音响度。

$$SPL_{2Pa}=20\lg\left(\frac{p}{p_{ref}}\right)=20\lg\left(\frac{2}{20\times10^{-6}}\right)=20\lg\left(1\times10^{5}\right)=100\text{dB}$$

声压增大 2 倍，则声压级增大 6dB；声压增大 10 倍，则声压级增大 20dB。

$$SPL_{2\mu Pa}=20\lg\left(\frac{p}{p_{ref}}\right)=20\lg\left(\frac{2\times10^{-6}}{20\times10^{-6}}\right)=20\lg\left(1\times10^{-1}\right)=-20\text{dB}$$

当实际声压小于参考声压时，声压级为负值。分贝的概念同样适用于声源的声强和声功率。

[注 1] 帕是压强的单位。1 帕（Pa）等于 1 牛顿每平方米（1N/m^2）。

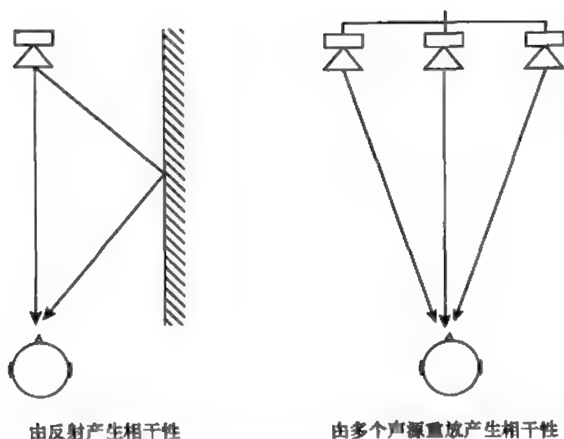
1.3 声波的叠加

目前为止，我们仅考虑了单个声源产生的声波大小。然而，在大多数情况下，几个声源会同时存在，这些声源可能来自不同的乐器，也可能来自房间的反射面。当进行声压叠加计算时，需要分下述两种情况考虑。

（1）相干声源：在这种情况下，声音来自相关联的几个声源。当其他几个声源是由同一个声源产生时，就属于这种情况。通常存在两种关联方式。其一，通过简单的反射产生其他声源，例如由声源附近的反射面反射形成。当延时较短时，延迟声就与原声非常近似，从而与原声产生很强的相干性。其二，声音来自同一个电子信号源，如录音节目、传声器等，并由多只扬声器重放。由于扬声器反馈给同一个信号源，并且在空间上存在一定的间距，因此可以认为是相关联的声源，具有相干性。这两种情形如图 1.8 所示。

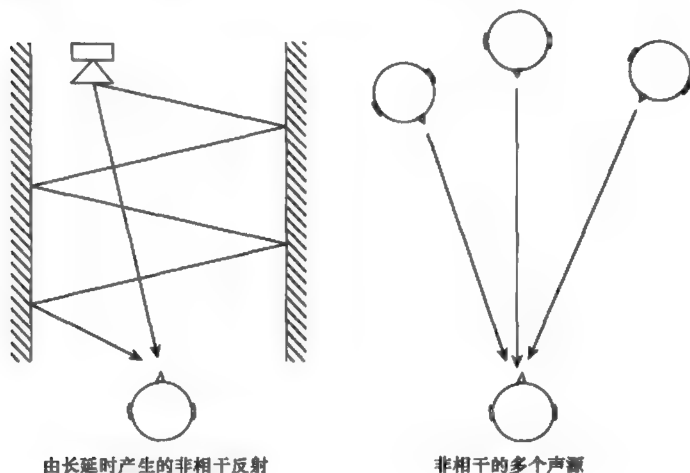
（2）非相干声源：在这种情况下，声音来自不相关联的几个声源。例如，来自不同乐器的声音，或者声音来自同一个声源，但由于反射产

图 1.8 相干声源的叠加



生的延时足够长，使得两个声音不具有相关性。对于第一种情况而言，不同的乐器将产生不同的波形和不同的频率成分，即使几个相同的乐器同时演奏，这种不一致性也会存在。对于第二种情况而言，另一个声源源于第一个声源，因此可能被认为具有一定的相关性，但是，由于延迟时间较长，就意味着第二个声源的波形将与原声源不同，这是因为在延迟一段时间后，原声源的音调、大小和波形已经发生了变化。由于延迟波与原声波没有了共性，因此可以认为与原声源无关，因此它们是非相干声源。这两种非相干声源的情形如图 1.9 所示。

图 1.9 非相干声源的叠加



1.3.1 相干声源的叠加

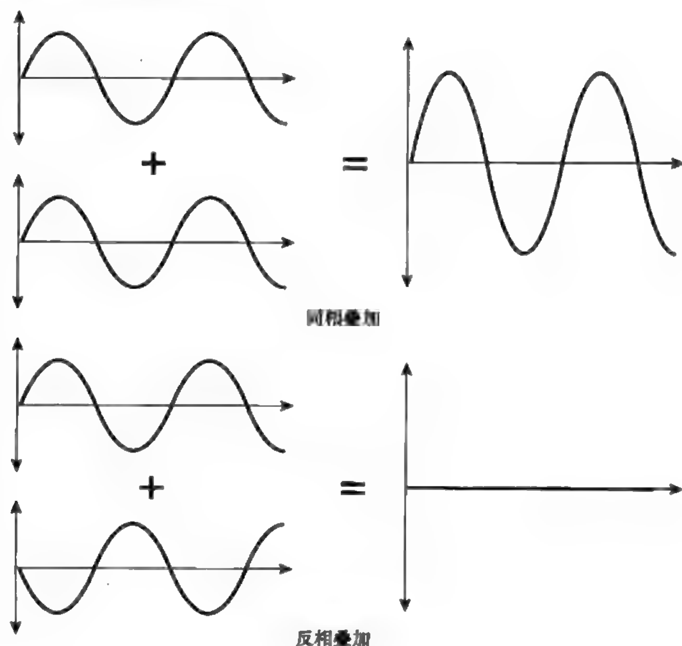
声级叠加时的计算方法，取决于声源是相干声源还是非相干声源。当声源是相干声源时，总声压波为各相干声源产生的声压波之和（译者注：指瞬时声压），如式（1.13）所示：

$$P(t)=P_1(t)+P_2(t)+\cdots+P_N(t) \tag{1.13}$$

值得注意的是，相干波一般具有相同的频率，彼此之间保持不变的时间关系，使得在某点叠加后的声压也是时间的函数。同时要注意到，由于叠加位置的不同会使相干波之间的时间关系发生变化，因此总声压与位置有关。

由于声波具有周期性（译者注：这里指正弦波），因此不同声源的声波由于相位不同而具有不同的正、负符号和振幅大小。例如，当两列等振幅波在某点为同相位时，其叠加结果是声压增大为单列波声压振幅的两倍；相反，当两列波在某点为反相关系时，由于相互抵消，使得该点的总声压振幅为零。这两种情况如图 1.10 所示。例 1.9 将说明单个反射声对某点声压振幅的影响。

图 1.10 不同相位正弦波的叠加



例 1.9 试计算声波在钢材和榉木中的传播速度。

某点的声音由来自扬声器的声音及其反射声组成，假设两个声音具有相同的振幅，反射声的延时为 1ms。试问在 250Hz、500Hz 和 1kHz 时该点的声压振幅分别是多少？

单一频率正弦波声压可用下式表示：

$$P=P_a\sin(2\pi ft) \quad 或 \quad P_a\sin(360°\ ft)$$

其中，f 为频率（Hz），t 为时间（s）。

函数中的系数 2π 或 360° 用来精确表示正弦波在一个周期中所处的位置。每个周期完成一周的转动，因此一个周期对应于转动 360°，或者更常见的说法是 2π 弧度。这种频率的表示方法称为角频率（1Hz=2π 弧度每秒）。

由于传播路径的变化引起的延时使其中一列波的到达时间发生改变。延时为 τ 的正弦波的声压表示为：

$$P_\tau=P_a\sin(2\pi f(t+\tau)) \text{ 或 } P_a\sin(360°\ f(t+\tau))$$

其中，τ 为延时（s）。

将延时波和未延时波叠加后得：

$$P_{total}=P_{a1}\sin(360°\ f(t+\tau))+P_{a2}\sin(360°\ ft)$$

如果延时波和未延时波等振幅，则

$$P_{total}=2P_a\cos\left(360°f\left(\frac{\tau}{2}\right)\right)\sin\left(360°f\left(t+\frac{\tau}{2}\right)\right)$$

上式中余弦项与延时和频率有关，而正弦项表示略有延时的原声波。因此，两列波叠加后的声压振幅为：

$$P_{a\,total}=2P_a\cos\left(360°f\left(\frac{\tau}{2}\right)\right)$$

用上式可以计算不同频率时延时波对总声压振幅的影响：

$$P_{a\,total\ 250Hz}=2P_a\cos\left(360°f\left(\frac{\tau}{2}\right)\right)$$

$$=2P_a\cos\left(360°\times250\times\left(\frac{1\times10^{-3}}{2}\right)\right)=1.41\,P_a$$

$$P_{\text{atotal 500Hz}} = 2P_a \cos\left(360°f\left(\frac{\tau}{2}\right)\right)$$

$$= 2P_a \cos\left(360° \times 500 \times \left(\frac{1\times10^{-3}}{2}\right)\right) = 0$$

$$P_{\text{atotal 1kHz}} = 2P_a \cos\left(360°f\left(\frac{\tau}{2}\right)\right)$$

$$= 2P_a \cos\left(360° \times 1\,000 \times \left(\frac{1\times10^{-3}}{2}\right)\right) = 2P_a$$

上述计算结果表明，相干声源叠加后的总声压与频率有很大关系，总声压振幅可以在（0 ～ 2）倍单个声压振幅值之间变化。

1.3.2 非相干声源的叠加

另一方面，如果声源是非相干的，除了可以像相干声源那样通过代数相加求得总声压的大小（译者注：指瞬时声压），它还满足总功率为每个声波的声功率之和。如前所述，声波的声功率与声压的平方成正比，所以要得到总声功率需要将声压有效值平方进行相加。总声压有效值为各声压有效值平方和的开平方根，用公式表示为：

$$P = \sqrt{\left(P_1^2 + P_2^2 + \cdots + P_N^2\right)} \tag{1.14}$$

非相干声源的叠加和相干声源的叠加在某些方面是不同的。首先，非相干声源的叠加只与所叠加信号的声功率有关，因此与声源的相对相位无关。这意味着非相干声源的叠加通常能提高声压级。第二个不同之处是，非相干声源叠加后声压级的增量没有相干声源大，因为叠加的是功率而不是声压。对于两个相同的相干声源，其叠加后最大的声压振幅可以是单个声压振幅的 2 倍；而两个非相干声源进行的是声功率的叠加，由于声功率与声压的平方成比例，这意味着非相干声源叠加后最大振幅仅为一个声源声压振幅的 $\sqrt{2}$ 倍。然而，非相干声源的叠加通常只会带来声级的增加，而不会产生相干声源叠加后可能带来的声压相抵消的现象。由于不存在相互抵消现象，非相干声源叠加后总声压随空间位置的变化要比相干声源小得多，因为非相干声源的叠加只与声源振幅有关。例 1.10

说明了几个等振幅非相干声源叠加的结果。

例 1.10 两个歌手以相同的声级进行演唱，试计算总声压相对于一个人演唱声压的大小。当 N 个歌手以相同声级合唱时，试计算合唱总声压相对于一个人演唱声压的大小。

多个非相干声源叠加后的总声压计算公式可由式（1.14）给出：

$$P=\sqrt{\left(P_1^2+P_2^2+\cdots+P_N^2\right)}$$

N 个等振幅的声源叠加后可简化为：

$$P_N=\sqrt{\left(P_1^2+P_2^2+\cdots+P_N^2\right)}=\sqrt{NP^2}=P\sqrt{N}$$

因此，等振幅的非相干声源叠加后总声压与声源数量的平方根成正比。如果只有两个声源，则：

$$P_2=P\sqrt{N}=P\sqrt{2}=1.41P$$

声源增加会对声压级（SPL）、声功率级（SWL）和声强级（SIL）带来怎样的影响？对于声功率级和声强级，由于是能量的累加，因此声源是相干声源还是非相干声源并不影响声功率级和声强级的结果。然而对于声压级而言，声源是相干声源还是非相干声源对结果是有影响的。在进行声源叠加计算时，什么情况选择什么声级度量标准是最需要考虑的问题。

1.3.3 分贝的相加

分贝是用对数标度来度量功率比的一种表示方法。因为对数的相加和乘积的对数是等效的，所以分贝的相加和声源振幅的叠加是截然不同的。显然对数的相加不是简单意义的相加求和。

要注意的是，当进行分贝相加时，需要将分贝值转换为原始的声压值，然后进行声压的求和计算，再将总声压转换为分贝值。例 1.11 将说明这个问题。

例 1.11 两个歌手分别以 69dB 和 71dB 的声压级进行演唱，试计算相对于一个人演唱时总声压级的增量。

由式（1.12）可得单个声源的声压级为：

$$SPL = 20\lg\left(\frac{p}{p_{ref}}\right)$$

对于多个非相干声源，声压级为：

$$SPL = 20\lg\left(\frac{\sqrt{P_1^2 + P_2^2 + \cdots + P_N^2}}{P_{ref}}\right) = 10\lg\left(\frac{P_1^2 + P_2^2 + \cdots + P_N^2}{P_{ref}^2}\right) \quad (1.15)$$

将声压级转化为声压得：

$$p^2 = 10^{\left(\frac{SPL}{10}\right)} P_{ref}^{\ 2}$$

其中，$P_{ref}^2 = 4\times10^{-10}\ \mathrm{N^2/m^4}$。

将声压级代入上式得：

$$P_{69\,dB}^2 = 10^{\left(\frac{69}{10}\right)} \times 4\times10^{-10} = 3.18\times10^{-3}\mathrm{N^2/m^4}$$

$$P_{71\,dB}^2 = 10^{\left(\frac{71}{10}\right)} \times 4\times10^{-10} = 5.04\times10^{-3}\mathrm{N^2/m^4}$$

将这两个声压值代入式（1.15）得：

$$SPL = 10\lg\left(\frac{P_{69\,dB}^2 + P_{71\,dB}^2}{P_{ref}^2}\right) = 10\lg\left(\frac{3.18\times10^{-3} + 5.04\times10^{-3}}{4\times10^{-10}}\right) = 73.1\mathrm{dB}$$

请注意，相加的总声压级只比两个声源中较响的声源大 2dB，而不是大 69dB。增大 69dB 是声源的分贝值直接相加的结果，所以两者是有很大差别的。

分贝值的相加等效于乘法运算这一性质在一些领域中得到很好的应用。当需要对一些量进行乘法运算时，就可以直接将其分贝值相加。换句话说，当所要进行的声级计算是乘法运算时，就可以直接将各自的分贝值相加。从这个意义上说，声级的分贝度量是非常有用的，因为在许多声学现象中，声波强度的变化是乘积关系，例如，声波透过墙壁传播时的衰减或壁面对声波的吸收。例 1.12 说明分贝相加在这方面的应用。

例 1.12 两个歌手分别以相同的声压级演唱，试计算总声压级（SPL）的增量，并计算 N 名歌手以相同的声压级演唱时总声压级（SPL）的增量。

非相干声源叠加的总声压为：

$$P_N = P\sqrt{N}$$

它可以表示为声压级的形式：

$$SPL_N = 20\lg\left(\frac{P\sqrt{N}}{P_{ref}}\right) = 20\lg\left(\frac{P}{P_{ref}}\right) + 20\lg(\sqrt{N})$$

该公式中，第一项表示的是一个声源的声压级，声源数量开平方根的分贝值表示的是多个声源叠加的声压级增量。因此这个公式可以改写为：

$$SPL_N = SPL_1 + 10\lg N$$

上述公式表示 N 个声压大小相等的非相干声源叠加的总声压级。例如，10 个声源发声，由于 10lg(10)=10，因此声压级会提高 10dB。

两个歌手演唱时的总声压级为：

$$SPL_2 = SPL_1 + 10\lg 2 = SPL_1 + 3\text{dB}$$

由此可以得出结论，两个大小相等的非相干声源叠加后声压级将提高 3dB。

1.4 平方反比定律

到现在为止，我们只讨论了声音在某一个方向上的传播。而实际上声音的传播是三维的，也就是说，声音的传播并不只是沿着某一个特定的路径，而是以声源为中心向空间各个方向辐射，如图 1.7 所示。

当声音从声源开始向外辐射时，声音随着距离的增大而变弱。这不是因为声音被吸收了，而是由于声能所覆盖的范围变大了。图 1.11 所示为一个球形的充气气球，外面覆盖了一定密度的蜂蜜，如果继续吹气使它的半径变为原来的两倍，它的表面积将会增至 4 倍，此时外表覆盖的蜂蜜的总量并没有变化，而密度却变为原来的 1/4。声强的变化和它是相似的，参考点与声源的距离每增大一倍，声强则衰减为原来的 1/4，

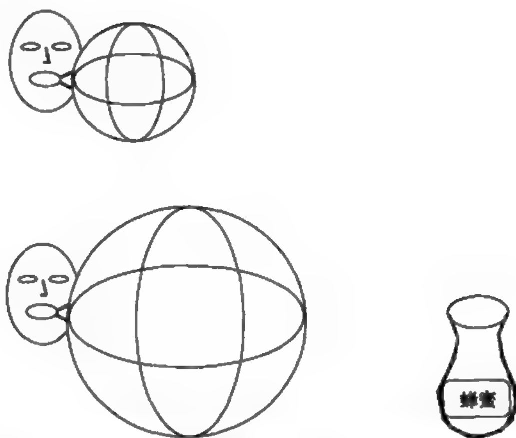


图 1.11　平方反比定律的蜂蜜气球模型

声强的大小和参考点与声源的距离的平方成反比。球体面积的计算公式为：

$$A_{sphere}=4\pi r^2$$

声强表示通过单位面积的声功率，声强为参考点到声源距离的函数可表示为：

$$I=\frac{W_{source}}{A_{sphere}}=\frac{W_{source}}{4\pi r^2} \tag{1.16}$$

其中，I 为声强（W/m^2），W_{source} 为声源的声功率（W），r 为参考点到声源的距离（m）。

式（1.16）表示了声波从声源向各个方向辐射时，声强随距离平方减小的规律。这种强度的衰减只是由距离引起的，不包含物理吸收的过程。实际上，空气声吸收的来源有很多，比如杂质分子和水分子的吸收，或者烟雾和湿气的吸收。这些声吸收在高频表现得较强，使得当人远离声源听音时，感到声音的响度减小，同时音色会变得暗淡。声能的这种额外衰减量取决于空气所含杂质的多少和湿度的大小，因此是一个变量。

例 1.13　一只扬声器的辐射声功率是 100mW，试问在距离扬声器 1m、2m、4m 处的声强级（SIL）分别为多少？这 3 个位置的声强级与扬声器的声功率级是什么关系？

声功率级可通过式（1.11）计算得出：

$$SWL = 10\lg\left(\frac{W}{W_{ref}}\right) = 10\lg\left(\frac{100\times10^{-3}}{1\times10^{-12}}\right)$$
$$= 10\lg\left(1\times10^{11}\right) = 110\ \text{dB}$$

在已知距离时，声强级可通过式（1.10）和式（1.16）计算得出：

$$SIL = 10\lg\left(\frac{I}{I_{ref}}\right) 10\lg = \left(\frac{\frac{W_{source}}{4\pi r^2}}{I_{ref}}\right)$$

上式简化为：

$$SIL = 10\lg\left(\frac{W_{source}}{W_{ref}}\right) - 10\lg(4\pi) - 10\lg(r^2)$$

可进一步简化为：

$$SIL = SWL - 20\lg r - 11 \qquad (1.17)$$

这个公式可用于计算上述 3 个距离的声强级，得：

$$SIL_{1m} = 110 - 20\lg1 - 11 = 110 - 0 - 11 = 99\text{dB}$$
$$SIL_{2m} = 110 - 20\lg2 - 11 = 110 - 6 - 11 = 93\text{dB}$$
$$SIL_{4m} = 110 - 20\lg4 - 11 = 110 - 12 - 11 = 87\text{dB}$$

由以上计算结果可知，距声源 1m 处的声强级比声源的声功率级小 11dB。请注意，理论上声源处的声强级是无穷大的，因为点声源的面积为零。而在实际应用中，所有的真实声源都具有一定面积，因此声强级的大小通常是有限的。我们还能看出这样一个规律，即声强级的大小随着距离每增大一倍衰减 6dB，这是通过平方反比定律直接推得的结果，也是一种简便的声级估算方法。图 1.12 所示为声源的声强级随距离的对数衰减的规律，它表明除了离声源很近的位置以外，距离每增大一倍、声强级衰减 6dB 的规律可以用一条直线表示。对于距离声源很近的情况，由于声源具有一定面积，所以在某种程度上说式（1.16）是无效的。近场区大致上是指声源的半径以内的区域。在此区域声场会发生急剧的起伏变化，取决于声源上各点振幅变化的具体情况。

式（1.16）说明了声源向各个方向辐射时声强的大小。然而，这个公式只适用于声源远离反射面的情况。这种没有反射面的声辐射，我们通常称之为自由场的声辐射，因为此时没有边界去阻碍声波的传播。

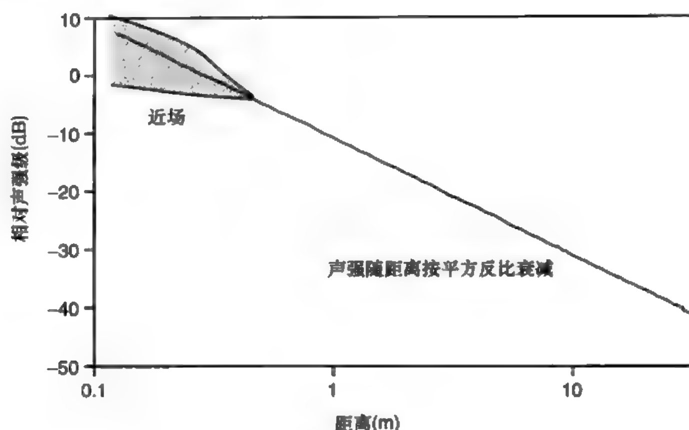


图 1.12 声强随到声源距离的变化特性

边界的影响

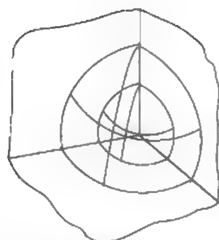


图 1.13 声源位于边界处的平方反比定律

然而在有边界的情况下，式（1.16）会有怎样的变化呢？显然，有很多声学文章都涉及了声源附近存在边界的情况，甚至在声源的四周都有边界存在的情况，例如房间里的声场，这部分内容将在第 6 章介绍。然而在很多时候，声源被放置在边界处，如放在地面上。在这种情况下，声音向一个有限的空间辐射，如图 1.13 所示。尽管辐射的空间是受限制的，然而波阵面的面积仍然和距离的平方成正比，正如图 1.13 所示。边界仅仅起到将声源的声功率集中在较小角度范围的作用。这个集中的效果可以用一个附加的乘法系数表示，因此声强计算公式可改写为：

$$I_d = \frac{QW_{source}}{4\pi r^2} \tag{1.18}$$

其中，I_d 为有指向性声源的声强（W/m^2），Q 为声源的指向性因数（和相同声功率的球面波的声强比），W_{source} 为声源的声功率（W），r 为到声源的距离（m）。

式（1.18）可应用于声源的辐射角度范围小于球面波的任何一种情况。显然边界的出现使辐射空间受到限制，但也有其他的方法能够达到同样的效果。例如，铜管乐器的号筒结构也可以改变其辐射的指向性。然而，需要注意的是，声强的减小仍然符合与距离的平方成反比的规律，而与

声源的指向性无关。

声源放置在某边界处的声强是可以计算的，下面举例说明边界对扬声器的辐射声强的影响。

例 1.14 某扬声器的辐射声功率是 100mW，将扬声器分别放置在相互垂直的 1 个、2 个、3 个边界处，试分别计算距离扬声器 2m 处的声强级。

已知到声源的距离，可以由式（1.10）和式（1.18）计算出声强级：

$$SIL = 10\lg\left(\frac{I}{I_{ref}}\right) = 10\lg\left(\frac{\frac{QW_{source}}{4\pi r^2}}{W_{ref}}\right)$$

上式可简化为：

$$SIL = 10\lg\left(\frac{W_{source}}{W_{ref}}\right) + 10\lg Q - 10\lg(4\pi) - 20\lg r$$

除了增加了一个指向性因数 Q 以外，这个公式和式（1.17）非常相似。相互垂直的 1 个、2 个、3 个边界将声波辐射的球面空间划分为半球面、1/4 球面和 1/8 球面，因此，它们仅在指向性因数 Q 上有区别。利用例 1.13 的计算结果，距离扬声器 2m 处的声强级分别为：

$$SIL_{1boundary} = SIL_{2m} + 10\lg Q = 93 + 10\lg 2 = 93 + 3 = 96\text{dB}$$
$$SIL_{2boundary} = SIL_{2m} + 10\lg Q = 93 + 10\lg 4 = 93 + 6 = 99\text{dB}$$
$$SIL_{3boundary} = SIL_{2m} + 10\lg Q = 93 + 10\lg 8 = 93 + 9 = 102\text{dB}$$

从计算结果可以看出，每增加 1 个边界声强可以增加 3dB，这是由于边界的出现使声辐射具有指向性。请注意，如果边界的数量超过 3 个，声音就会碰撞到其他边界产生反射，这时就不能使用上述公式了。这个问题将在第 6 章进行阐述。为了使声源的指向性因数大于 8，可以通过其他技术手段，例如，可以方便地从扬声器制造商获得指向性因数为 50 的号筒扬声器。

1.5 声波的反射、干涉和衍射

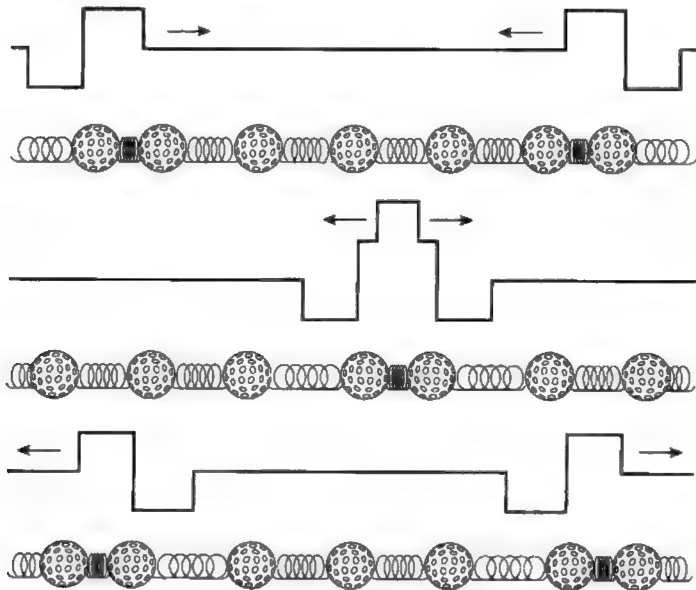
前面内容只涉及单个声源，并介绍了声速、频率、波长和平方反比定律等概念。然而，声音还会与障碍物或其他声波之间产生相互作用，

同时传播媒质的改变也会对声波产生影响。这一节将讨论声音的反射、干涉和衍射现象，这对于研究声学乐器的工作原理以及建筑中声学传播过程是非常必要的。

1.5.1 叠加原理

当两列声波反相相交时，声波并不会消失，而是穿过彼此沿着原来的路径继续传播下去；当两列声波同相相交时，它们也不会变强，只是穿过彼此而已。这是因为虽然声波的总声压或总振速会在零和两列波的声压和或振速和之间变化，但声波的能量能够得以保持并沿着传播方向继续传播下去。因此空间中某位置的总声压或总振速是各列声波传播到该位置的声压或振速之和，如图 1.14 所示。声波的这个特性称为线性叠加原理。由于声波满足线性叠加原理，这对我们描述和分析空间中某位置的声波特性是非常有帮助的。

图 1.14 高尔夫球弹簧模型中声波的叠加



1.5.2 声波的折射

声波的折射和光在不同媒质界面发生折射的现象很相似。光折射是由

于在不同媒质中光速不同而产生的，例如水中的光速小于空气中的光速。声波折射的原因也是相似的。空气中的声速与温度有关，如式（1.5）所示。

对于图 1.15 所示的两种不同温度的空气界面，根据斯奈尔定律，当声波以一定角度入射到分界面时，声波的传播方向就会改变，由式（1.5）得：

$$\frac{\sin\theta_1}{\sin\theta_2}=\frac{v_{T1}}{v_{T2}}=\frac{20.1\sqrt{T_1}}{20.1\sqrt{T_2}}=\sqrt{\frac{T_1}{T_2}} \tag{1.19}$$

其中，θ_1、θ_2 为声波传播方向与两种媒质分界面法线的夹角，v_{T1}、v_{T2} 为声波在两种媒质中的传播速度，T_1、T_2 为两种媒质的绝对温度。

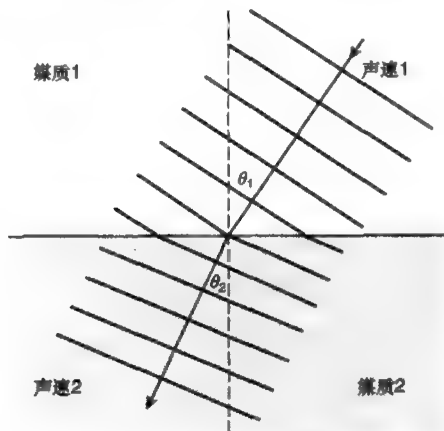


图 1.15 声波的折射（媒质 1 的绝对温度是 T_1，声速是 v_{T1}；媒质 2 的绝对温度是 T_2，声速是 v_{T2}）

因此，声波的传播方向是随两种媒质的绝对温度比的平方根而变化的。声速随着温度的升高而增大，当声波从冷空气传播到热空气时，声波的折射角度将变大，而从热空气传播到冷空气时，声波的折射角度将减小。这个特性可以用来解释一些户外的声传播现象。通常空气的温度随着高度的增大而降低，因此，当声波离开声源向远处传播时，声波传播方向呈向上弯曲状态，如图 1.16 所示。这意味着当地面上的听音者离声源一定距离听音时，听到的声压会有所衰减，而且比平方反比定律衰减得快。这种现象有利于减小噪声对人的干扰。然而，如果温度随着高度的增加而增大，则声波不会向上弯曲而是向下弯曲，如图 1.17 所示。这种现象经常发生在夏季的夜里，在一定距离处的声压会大于平方反比定律的计算结果。在夏季晚间开露天流行音乐会时，远距离的人们会受

到较大的噪声干扰，而离音乐会地点较近的人们反而不会受到这么大的干扰，这种现象正是由声波的折射引起的。

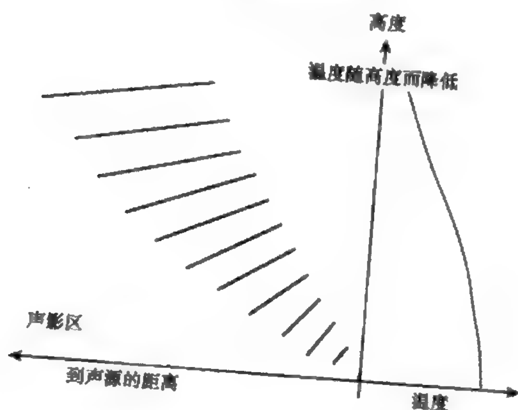


图 1.16 由垂直方向的温度梯度产生的声波折射现象

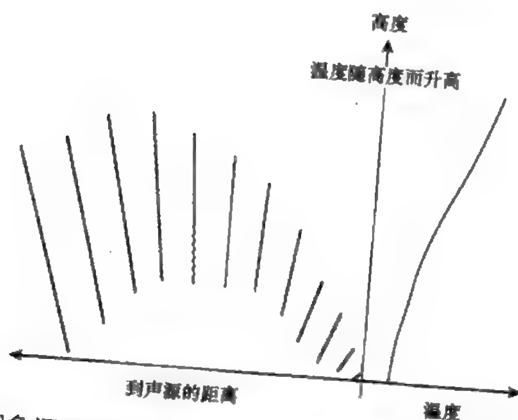


图 1.17 由相反的温度梯度产生的声波折射现象

折射现象还可能发生在两种不同温度的液体如水的交界面处，在某些时候折射甚至可能引发全反射。潜水艇有时用这种效应来躲避其他船舶的声呐。折射还可以使声波在两种边界之间进行传播，在这种情况下，声音可以传输较远距离。人们认为鲸鱼和海豚就是利用这个原理在海洋中用声音进行长距离沟通。

风也可能引起折射效应。由于声速不随介质速度的变化而变化，因此在运动的介质中，对于某一固定位置，声速是两者速度之和。因此，当声音传播的方向和风速方向相同时，声速增大；当声音传播的方向和风速方向相反时，声速减小。地面处空气的速度一般比地面上方空气的

速度小（由于地面处受到摩擦力的作用），声波传播方向向上或向下弯曲取决于声波和风的相对方向。声波方向的改变量取决于风速随高度的变化率，风速的变化率越大则声波方向的改变量就越大。图 1.18 所示为风对声音传播的影响。

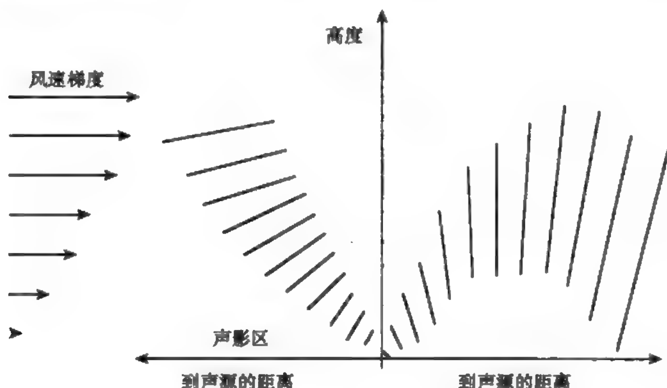


图 1.18　声波由于风速梯度产生的折射现象

1.5.3　声吸收

当声波作用在物体上时，会发生声吸收。原因之一是当声波碰撞到一个物体时，如果物体不是刚性体，就会产生振动。这就是说声能转换成物体振动的能量，然后物体内部材料的摩擦损耗会吸收一部分能量。声波在撞击、穿过一些多孔材料时也会损耗一部分能量。这是由于这些材料有很多的纤维和孔洞，因此和声波有很大的接触面，使声波的速度分量与物体表面相互作用时产生摩擦损耗。接触面越大，能量的损耗越大，因此多孔材料（如织物或岩棉）具有很强的吸声能力。

1.5.4　刚性边界的声反射

当声音碰撞到刚性边界时，也会发生反射。我们都感受过声反射。例如，当我们靠近一个面积很大的刚性物体如悬崖或大型建筑时，能够听到回声。声音的反射主要包括以下两种情况。

第一种情况是，声波撞击某个固定的物体或刚性边界，如图 1.19 所

示。由于声波不能推动墙面，因此在物体和空气之间的分界面上声速为零。这就是说，在这个位置上，声波的所有能量都压缩在空气中或体现在声压中。由于存储在声压中的能量不能沿传输方向继续传播，因此它会沿着相反方向反射，从而改变了声波速度分量的相位。图 1.19 所示的高尔夫球弹簧模型说明了这种声波反射现象。由于声波需要改变传播方向，连接固定边界端的弹簧压缩量变为原来的两倍。这种情况是由于反射波速度分量的相位被改变，从而使反射波的速度分量与入射波速度分量的相位相反造成的。从声学角度上说，当发生反射的边界上速度分量为零时，压力分量变成原来的两倍。

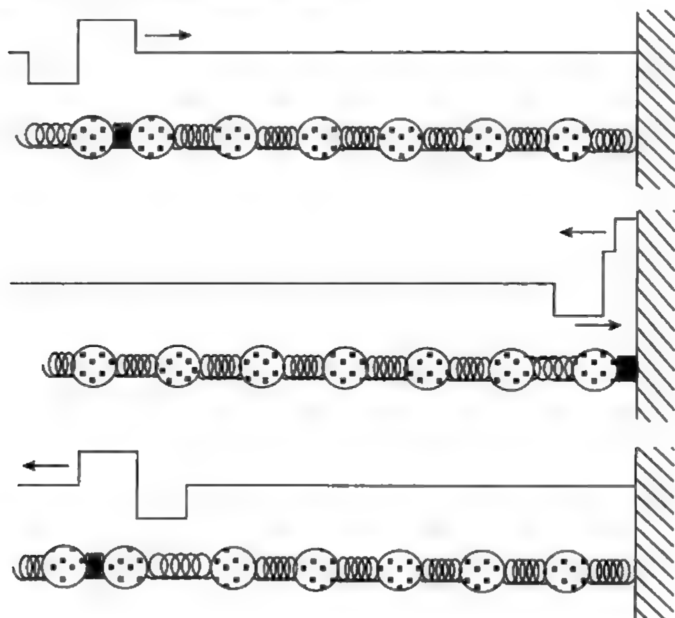


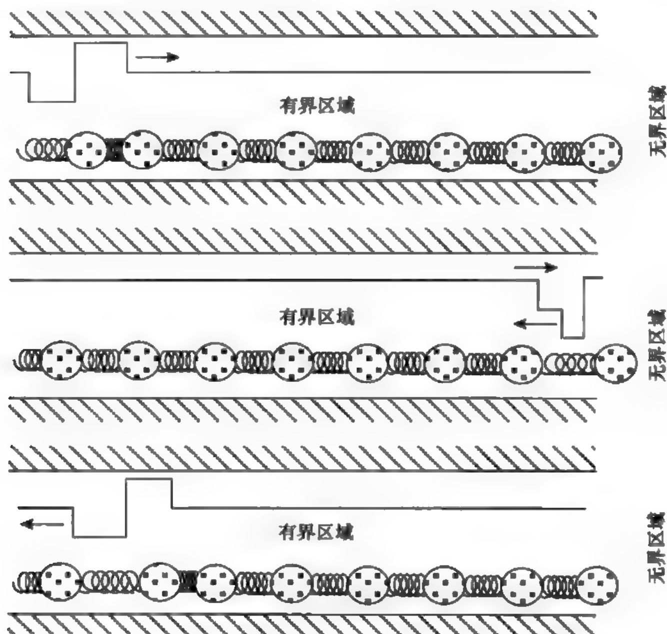
图 1.19 声波遇到刚性界面的反射现象

1.5.5 从有界区域到无界区域的声反射

第二种声反射情况是，声波从有界区域传播到无界区域。例如，声波从管子传播到无界区域即自由空间 [注 1]，如图 1.20 所示。在有界区域和无界区域的分界处，无界区域的空气分子比有界区域的空气分子容易移动得多。其结果是，在分界处声波有一个接近零的声压和很大的速

度。因此，这个位置声音的所有能量全部转化为空气分子的动能。换句话说，转化为振动速度。由于在无界区域运动阻力较小，即“弹性”较小，所以储存在速度分量的能量不能沿着运动方向继续传播。因此，质点在分界处被进一步向前推伸后，分子的动量迅速返回到有“弹性”的有界区域。这相当于一个传播方向相反的反射波，反射波的相位是相反的。因为它是从弹簧的拉伸状态或稀疏状态开始传播的，而不是从压缩状态开始。图 1.20 所示为用高尔夫球弹簧模型说明这种现象，其中无界区域用没有任何弹簧来模拟。在这种情况下，由于声波需要改变传播方向，发生了一个有趣的现象，即连接无界区域的质量块以两倍于离分界面较远的质量块的速度运动。产生这种现象的原因是，由于反射波压力分量的相位改变了，使得反射波的压力分量与入射波的压力分量相位相反。从声学意义上说，发生反射的边界处的声压分量为零，而速度分量是原来的两倍。

图 1.20　声波从有界区域到无界区域的反射现象



综上所述，在硬边界发生反射后，反射波与入射波的压力分量相位相同，然而从有界区域到无界区域发生反射后，反射波与入射波的压力分量反相。这是由于不同边界的声阻抗不同所导致的。在第一种情况下，

边界的阻抗远大于传播媒质的阻抗；在第二种情况下，边界的阻抗远小于传播媒质的阻抗。当声波斜入射到边界时，同样可以应用反射定律。

[注 1] 自由空间是声波可以自由地在任何方向传播的区域。在这个区域里没有障碍物，传播媒质也是均匀的，因此声波的传播不受任何影响，并且自由空间是一种无界区域。然而不是所有的无界区域都是自由空间，例如声波从很大的墙面上的管中传播出来，此时虽然存在一个从有界区域到无界区域的变换，但这不是自由空间，因为声波不能往任何方向传播。

1.5.6 声波的干涉

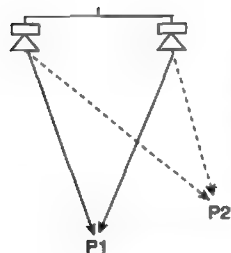


图 1.21 相干声源的干涉

之前我们介绍过，当声音来自几个相干声源时，其声压和速度是各声压分量和速度分量的简单相加。这就是说声压幅度可能在零和两倍声压幅度之间变化，如图 1.10 所示。两列声波叠加后的结果是增强还是减弱，取决于它们的初始相位以及每列声波的传播距离。由于声波在空间中的分布根据波长的不同而各异，因此相位在空间中的分布也是不同的。这就是说声波叠加后是增强还是减弱，与空间位置有关。在图 1.21 中，两个相干声源在室内进行重放，当听音者位于 P1 位置时，由于到两声源的距离相等，因此两列声波是同相的，所以叠加的结果声压是增强的；当听音者位于 P2 位置时，由于到两声源距离不相等，因此两列声波不一定是同相的。实际上，如果两列声波的声程差为半波长，则这两列声波将反相抵消。以上现象称为声波的干涉，即相干声源之间发生了干涉现象。请注意，非相干声源不会发生干涉现象。声波的相对相位取决于声源的声程差或相对延时量。正因如此，是相长干涉还是相消干涉在很大程度上取决于参考点的位置，如图 1.22 所示，并且与声波的频率有关，因为声源和参考点相对于波长（λ）的距离决定了声波是发生相长干涉还是相消干涉。对于特定位置，声压幅度频率特性的形状看似梳状的齿一样，这种干涉引起的频率特性的变化通常称为“梳状滤波（comb filtering）”效应。由于波长与频率成反比，因此干涉效果会随着频率变化而变化，情况确实如此。图 1.23 所示为两列距离不同的等振幅声波叠加后的声压振幅。叠加后的振幅大小随位置（用距离差表示）而变化，图中距离差用波长（λ）的倍数表示。图 1.23 说明，当两列声波的距离差为

声波波长的整数倍时，叠加后幅度会增强；当声波的距离差为半波长的奇数倍时，叠加后幅度会抵消。对于特定的位置，距离差所对应的波长数量随频率的变化而变化，因此特定位置声波的干涉效果也随着频率的变化而变化，其频率特性如图 1.23 所示。如果两列波幅度不等，干涉的效果就会减弱，见图 1.23。实际上，如果一列声波的振幅小于另一列声波的 1/8，则叠加后声压级的峰值波动范围将小于 1dB。

图 1.22 频率一定时声波在不同位置的干涉现象

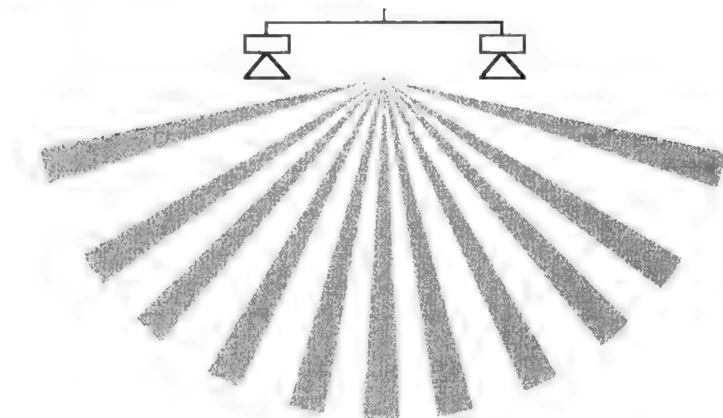
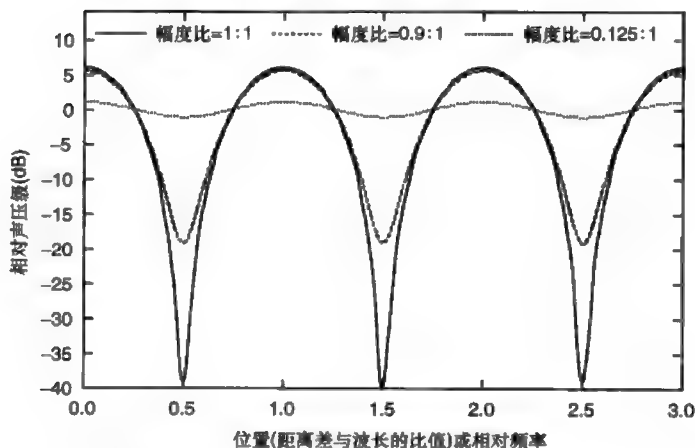


图 1.23 特定位置声波干涉引起的声压随频率变化特性（比值代表两列波的振幅比）



在一些声学情况下，可能引起干涉现象。最常见的是，当两个扬声器在一个房间辐射同一个声源时，或者当同一个声源通过两个不同的开口耦合到同一个房间时。其他发生干涉的情况是，单个声源以一定距离放置在反射面的一侧。在此情况下，由反射声产生的镜像声源与原声源

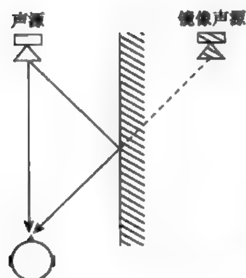


图 1.24 由壁面反射引起的干涉现象

可形成相干声源，如图 1.24 所示。这种情况经常出现在录音或扩声中，由于传声器同时拾取了直达声和反射声，因此会产生声波干涉的问题。

例 1.15 两只扬声器间距 1m，辐射相同的声压级。听音者正对其中一只扬声器并距离 2m 听音，并与两扬声器连线成直角，如图 1.25 所示。试问声波发生相消干涉的两个最小频率分别是多少？除了频率很低的情况外，声波发生相长干涉的最小频率是多少？

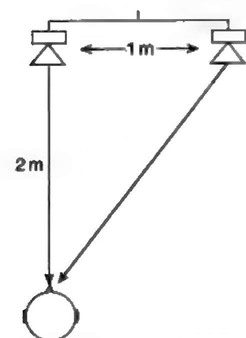


图 1.25 两只扬声器在某点发生干涉现象

通过勾股定理首先算出两列声波的声程差：

$$\Delta_{path}=\sqrt{(1^2+2^2)}-2=0.24\text{m}$$

声波发生相消干涉时，声程差应等于 $\lambda/2$ 和 $3\lambda/2$。因此，第一个频率可以用下式计算：

$$\frac{\lambda}{2}=\Delta_{path}$$

$$f_{\lambda/2}=\frac{v}{\lambda}=\frac{v}{2\Delta_{path}}=\frac{344}{2\times0.24}=717\text{Hz}$$

第二个频率可以用第一个频率的 3 倍来计算：

$$f_{3\lambda/2}=3\times f_{\lambda/2}=3\times717=2\ 150\text{Hz}$$

发生相长干涉的最小频率可以用相消干涉的最小频率的 2 倍来计算：

$$f_{\lambda}=2\times f_{\lambda/2}=2\times717=1\ 434\text{Hz}$$

如果听音者向两扬声器的中线方向移动，则扬声器间的延时差会减小，相消干涉的最低频率将会提高。在极限情况下，也就是当听音者距离两扬声器相等时，声波不会发生抵消，也就是相消干涉的频率是无限大。

1.5.7 刚性边界内的驻波（共振模式）

无论声波是以什么样的形式存在，声波的线性叠加原理也可以用于解释一种称为驻波的波动现象。当声波在反射面之间来回反射时，就会形成驻波。图 1.26 所示为一个包含两个反射面的最简单的产生驻波的结构。在此结构中，声波在反射面之间往返传播。在大部分频率下，反射

面之间的距离与声波的波长不成整数比例，因此波峰和波谷可能出现在界面间的任何位置，并且出现的概率相等，如图 1.27 所示。

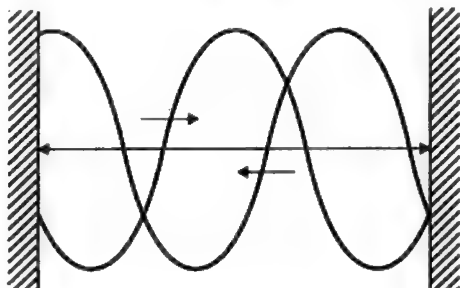


图 1.26 声波在两个平行界面间的反射

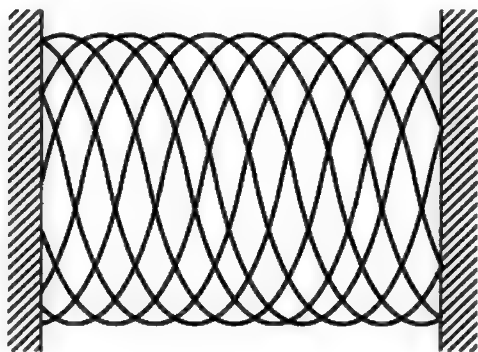


图 1.27 两个平行界面间传播的疏密位置不固定的声波

然而，当波长和反射面的间距成一定整数比例时，声波将沿着相同的轨迹在反射面之间往返传播，即声波的疏密状态在两个反射面之间的位置是固定的。因此，声波在反射面之间的状态是固定不变的，因而称为驻波，更确切地说，是一种共振模式。[注 1] 驻波实际上是以一个正常的速度传播的，只是两列波叠加后波形并不向前推进，如图 1.28 所示。从图 1.28 和图 1.29 可以看出驻波在两个刚性界面间的声压分量和速度分量的分布情况，此时，在两个反射面处声压分量达到最大值，速度分量达到最小值。满足以上条件的最低频率是半波长等于反射面的间距，因此在一定间距的两个反射面之间可能存在的驻波最低频率可用下式计算：

$$f_{\text{lowest}} = \frac{v}{\lambda} = \frac{v}{2L}$$

其中，f_{lowest} 为驻波的最低频率（Hz），L 为反射面的间距（m），λ 为波长（m），v 为声速（m/s）。

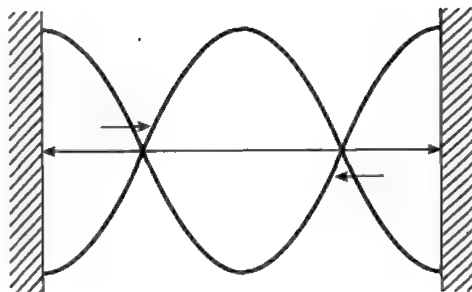


图 1.28 两个刚性界面中驻波的声压分量（也称为共振模式）

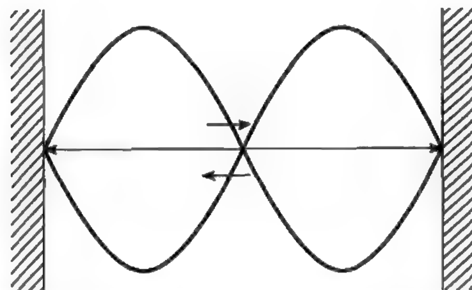


图 1.29 两个刚性界面中驻波的速度分量

当反射面的间距等于半波长的倍数时也能形成驻波，因此理论上形成驻波的频率有无数个，这些频率都是 f_{lowest} 的整数倍，并可以用下式计算：

$$f_n = \frac{nv}{2L} \tag{1.20}$$

其中，f_n 为第 n 次驻波频率（Hz），$n=1, 2, 3, \cdots, \infty$。

从图 1.28 和图 1.29 可以看出，驻波的声压分量和速度分量在某些位置产生最大幅度，在某些位置产生最小幅度。例如，在图 1.28 中，声压分量的幅度在边界处和中点处达到最大值；而在图 1.29 中，速度分量在边界处和中点处为零。声压分量幅度为零的位置称为压力波节点，声压分量幅度最大的位置称为压力波腹点。随着反射面之间驻波的半波长数量的增多，波节和波腹的数量随之增多。刚性界面压力波节点的数量等于半波长的个数，压力波腹点的数量比节点多一个。同理也存在速度波的节点和腹点，它们与压力波节点和压力波腹点是密切相关的，速度波腹点和压力波节点会同时出现，反之亦然，如图 1.30 所示。这是因为能量在声波的传输过程中，或者存储在压力波节点的速度分量中，或者存

储在速度波节点的声压分量中。

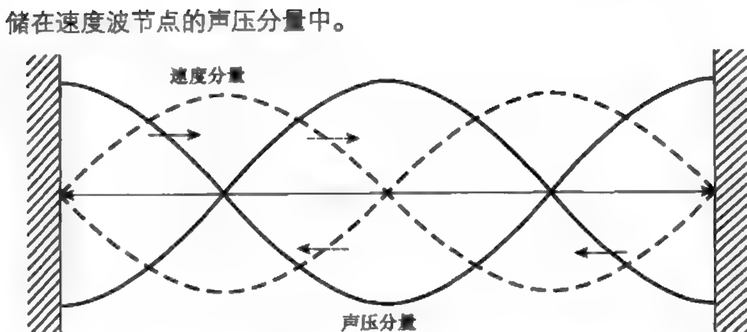


图 1.30 两个刚性界面中驻波的声压分量和速度分量

[注 1] 严格地说我们这里讨论的驻波应称为共振模式，因为无需利用两个边界也可以产生驻波。例如，我们不需要边界仅用两只一定间隔的扬声器即可产生任意频率的驻波，或者可以用一只扬声器或一个声源和一个边界来产生驻波。

1.5.8 其他边界中的驻波

驻波除了可以在上述刚性反射面之间传播外，还可以在以下两种边界条件下传播。第一种情况如图 1.31 和图 1.32 所示，声波从有界区域传输到无界区域，相当于在两端存在边界，例如两端开放的管子。在这种情况下，在边界处的声压为零，而速度达到最大值，如图 1.31 和图 1.32 所示。和刚性界面中的驻波类似，驻波的最小频率应满足半波长等于边界间距的条件，其他驻波频率为最小频率的整数倍。因此式（1.20）同样适用于计算这种情况下的驻波频率。

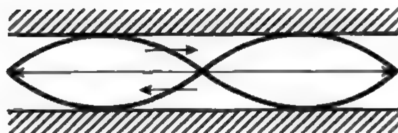


图 1.31 两个有界到无界边界中驻波的声压分布

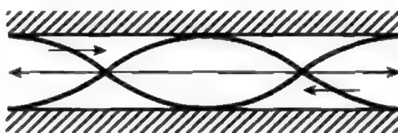


图 1.32 两个有界到无界边界中驻波的速度分布

例 1.16 有一根管长为 98.5cm 的两端开放的管，试计算最小的两个共振模式（驻波）频率。

由式（1.20）可得：

$$f_1=\frac{nv}{2L}=\frac{1\times344}{2\times0.985}=174.6\text{Hz}$$

$$f_2=\frac{nv}{2L}=\frac{2\times344}{2\times0.985}=349.2\text{Hz}$$

这两个频率与音阶中音符 F3 和 F4 相对应，它们之间满足倍频程的关系。

第二种情况是声波在一端为刚性边界另一端开放的管子中传播，如图 1.33 和图 1.34 所示。在这种情况下，开放端产生了压力波节点，刚性边界处产生了压力波腹点。在这种边界条件下，驻波存在的条件是两个边界之间存在奇数个$\frac{1}{4}\lambda$。$\frac{1}{4}\lambda$的偶数倍不能形成驻波，因为此时正如图 1.28 和图 1.31 所示会在两端同时生成一个压力波节点或压力波腹点，不能满足边界条件。因此，驻波应满足的条件是，其频率为最小驻波频率的奇数倍，用代数式可表示为：

$$f_n=\frac{(2n+1)v}{4L} \tag{1.21}$$

其中，f_n 为第 n 次驻波的频率，n=0，1，2，…，∞。

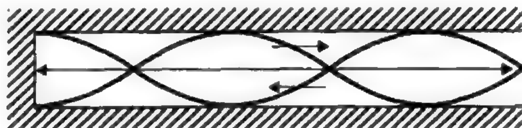


图 1.33　两种不同边界中驻波的声压分布

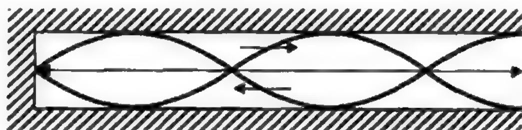


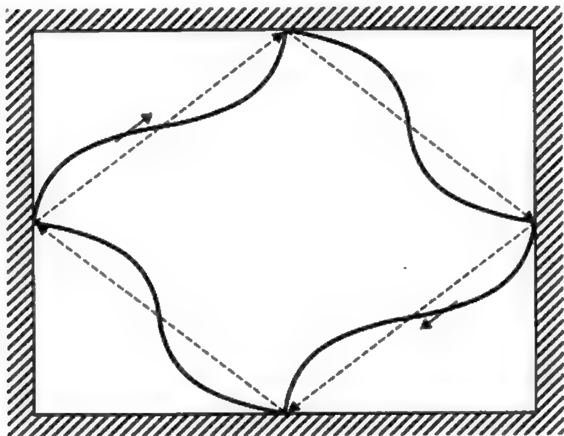
图 1.34　两种不同边界中驻波的速度分布

图 1.33 和图 1.34 是 n=2 的驻波，该驻波的频率是最低频率的 5 倍。驻波可以存在于任何形式的声波传播中。假设弦上横波的传播速度为 v，则两端固定的弦上的横波形成的驻波频率也可以用式（1.20）来计算。

驻波在声学中通常称为特定系统的模式。驻波的最低频率称为第一次模式，最低频率的倍数为高次模式，因此系统的第三次模式是驻波频率第三低的模式。产生驻波的边界条件也不仅限于两个平行的反射面，实际上任何反射或折射回来的声波均能产生驻波或模式。驻波可以以任

何形式存在于一维、二维或三维空间中。产生驻波的必要条件是声波的传输路径是周期性重复的，从而保证声波的每一次传输是同相位的。图 1.35 所示为二维驻波的一个例子。

图 1.35 二维驻波



例 1.17 如果将例 1.16 中的管子一端封闭，试计算前两个模式（驻波）的频率。

由于边界条件不同于例 1.16，因此要用式（1.21）计算驻波频率：

$$f_1=\frac{(2n+1)v}{4L}=\frac{(2\times0+1)\times344}{4\times0.985}=87.3\text{Hz}$$

$$f_2=\frac{(2n+1)v}{4L}=\frac{(2\times1+1)\times344}{4\times0.985}=261.9\text{Hz}$$

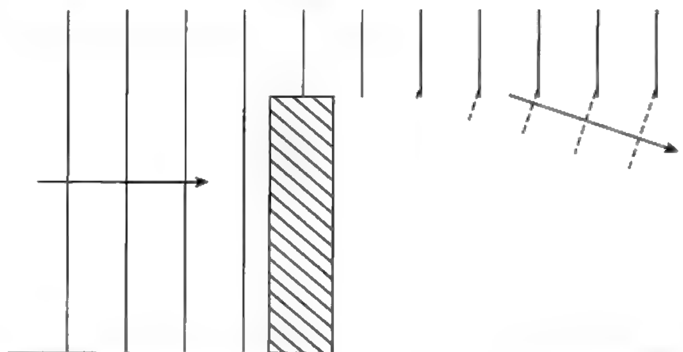
该例题中的第一次模式频率是两端开放管频率 174.6Hz 的一半，在音乐的音阶中这个简正频率比 F2 低一个八度。第二个模式频率是最低模式频率的 3 倍，这个频率与音阶中 C4 的频率大致相同。

1.5.9 声波的衍射

众所周知，声音具有绕过建筑物或障碍物的能力。在声波传播过程中，偏离直线的传播方向而绕过障碍物继续向前传播的现象，叫做声波的衍射（也称为绕射），如图 1.36 所示。当声波传播到障碍物边界时，由

声波的疏密变化导致的空气压强变化不能突然减小为零，这是因为传播声波的相邻分子之间有相互作用力。为了使声波的疏密变化从边界处到阴影区逐渐消失，有一部分区域的声波传播方向会发生改变，因此这部分改变传播方向的声波发生了绕射。

图 1.36 声波遇到障碍物的绕射现象



声波绕射的程度与波长有关，因为只有一部分的波长能够传播到阴影处。频率越低，声波越容易在建筑物或墙等边界处发生显著的绕射现象，如图 1.37 和图 1.38 所示。当声波穿过小孔时也会产生类似的现象，如图 1.39 和图 1.40 所示，这时声波沿着离开孔洞边缘的方向偏离原来的传播方向。绕射的程度取决于小孔相对于波长的大小，当波长相对于小孔尺寸较大时声波的绕射现象比较明显，当波长相对于小孔尺寸较小时绕射现象比较弱。波长较大或较小通常是以小孔孔径为波长的三分之二（$2/3\lambda$）为分界线。

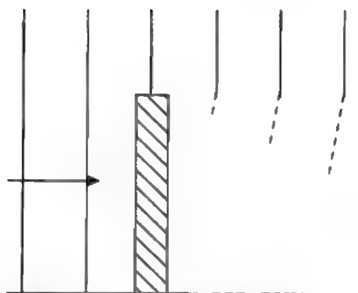


图 1.37 低频时声波遇到障碍物的绕射

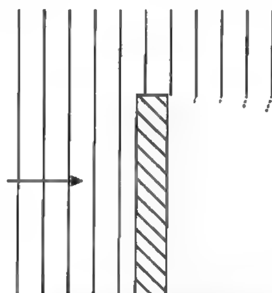


图 1.38 高频时声波遇到障碍物的绕射

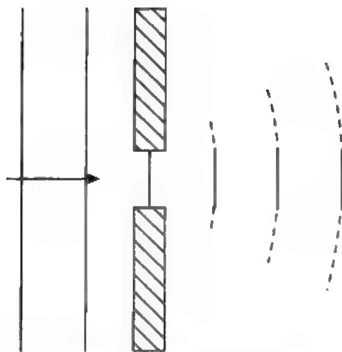


图 1.39　低频时声波穿过小孔的绕射

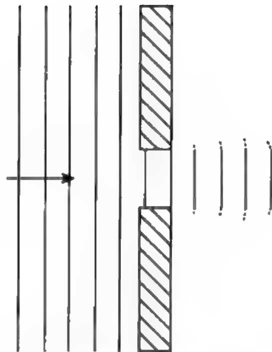


图 1.40　高频时声波穿过小孔的绕射

与声波穿过小孔的情况类似，当声波传播到固体障碍物时也会发生绕射，如图 1.41 和图 1.42 所示。我们可以把它们看作是孔洞的相反情况，此时声波绕射到障碍物后并且相遇。绕射的程度也是由障碍物相对于波长的尺寸决定的。当波长相对于障碍物的尺寸较大时声波的绕射现象比较强；当波长相对于障碍物的尺寸较短时绕射现象比较弱，这时会在障碍物背后产生较为明显的阴影区，如图 1.42 所示。障碍物的存在对声波的影响是否显著通常是以障碍物的尺寸为波长的三分之二（2/3λ）为分界线。

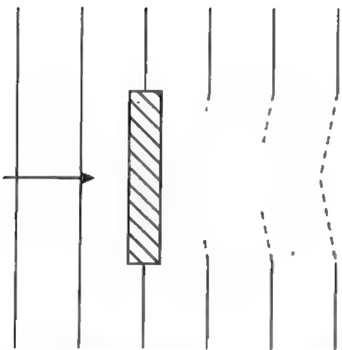


图 1.41　低频时声波遇到障碍物的绕射

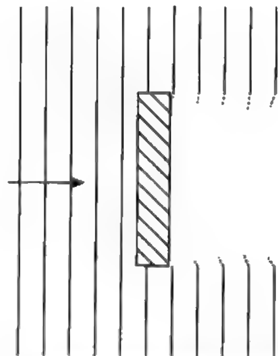
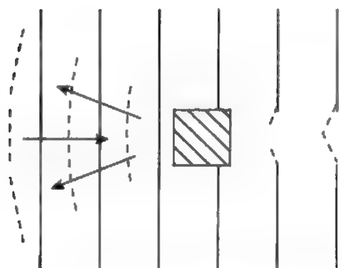
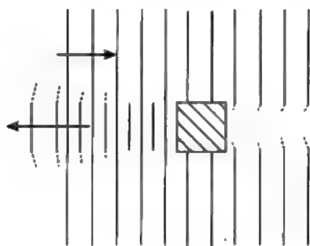


图 1.42　高频时声波遇到障碍物的绕射

1.5.10　声波的散射

入射到物体表面的声波不仅发生衍射，部分能量将被反射或散射，如图 1.43 所示。和衍射的情况类似，物体相对于波长的大小决定了散射能量的多少。当物体相对于波长较大时，根据反射定律，大部分的入射声能将被散射，只有少量的入射声能发生衍射，如图 1.44 所示。然而，当物体相对于波长较小时，只会有少量的入射声能被散射，大量的入射声能发生衍射，如图 1.43 所示。通常以障碍物的尺寸为波长的三分之二（2/3λ）做为衡量衍射强弱的基准。当物体尺寸小于这个基准时，散射往往是全方位的；当物体的尺寸大于这个基准时，散射的方向性变强。


图 1.43　低频时声波遇到障碍物的散射


图 1.44　高频时声波遇到障碍物的散射

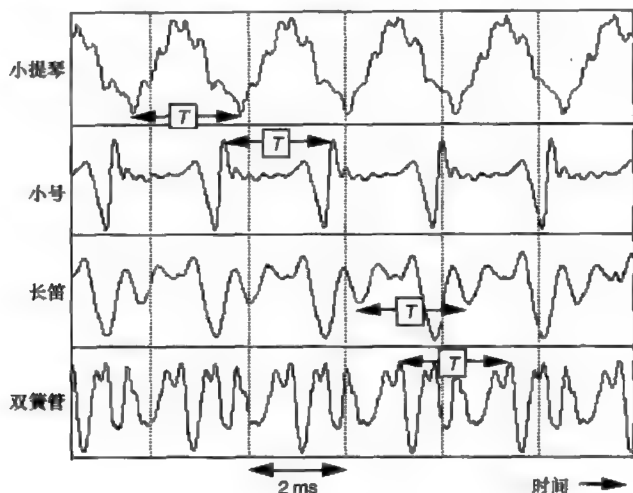
当引起散射的障碍物尺寸约为波长的三分之二（2/3λ）时，除了发生散射，还会发生干涉。这是由于在这个频率下还会发生反射和衍射。在这种情况下，障碍物两侧的声波会相互影响而发生干涉，使得在某些特定的频率上障碍物附近的反射或衍射效果得到加强。具体的干涉效果取决于发生散射的障碍物前后之间的相互作用，因此与障碍物的形状有很大关系。所以，对于球面体和矩形板而言散射的效果是不同的。由于衍射和散射的能量总和是一个常量，所以反射和衍射的频率响应在很大程度上是互补的。

1.6　时域和频域

到目前为止，我们考虑的是单频正弦声波的情况，这对描述与波长有关的传播现象非常有帮助。然而，大多数的音乐具有更加复杂的波形，

它们不只是一个简单的正弦波，如图 1.45 所示。那么，我们如何能够分析真实音乐的波形、如何用声学术语去描述它们呢？我们可以用声波叠加原理和傅里叶分析方法解决上述问题。

图 1.45　几种乐器声音的波形



1.6.1　周期性声波的频谱特性

利用傅里叶分析方法，任何一个随时间作周期性变化的波，都可以分解为一系列不同频率、不同振幅和不同相位的正弦波。图 1.46 所示为 4 个不同频率的正弦波叠加形成的波形。4 个正弦波的频率分别为 $1F$、$3F$、$5F$ 和 $7F$，相互之间没有相位差（它们从同一时刻开始振动，如图中虚线所示），并且幅度与频率成反比。这就是说，$3F$ 是 $1F$ 幅度的 1/3，以此类推。这些正弦波叠加到一起后形成的波形接近于方波，如图 1.46 所示。如果更多的高频成分叠加进去，叠加后的波形将更趋近于一个理想方波。高频成分可以让叠加形成的方波具有陡峭的上升前沿和尖锐的转角。一般情况下，如果希望叠加的波形具有更加快速的上升时间和更加尖锐的转角，需要更多高频正弦波去准确地描述波形。换言之，我们可以把方波看成很多正弦波叠加的结果，这些正弦波的频率都是基频的奇数倍，并且其幅度和频率成反比关系。一个正弦波代表一个频率，因此，给定幅度的正弦波可以用一根线段表示幅度与频率的关系，并描绘在坐标

图中。方波的各频率成分如图 1.47 所示。由图可知，方波是由一系列离散的正弦波叠加而成，这些频率都是最低频率的奇数倍，这种波形的表示方法叫做频域表示法或波形的频谱。波形的幅度随时间变化的表示法为时域表示法。波形的每一个正弦波分量可称为波形的分音。如果分量频率之间存在倍数关系，像上述方波那样，这些分量称为谐波。最低频率的分量称为基波，也可称为一次谐波，更高频率的谐波分量用其与基波之间频率的倍数关系来命名，因此，二次谐波的频率是基频的两倍，以此类推。另一方面，分音不一定与基波频率成整数倍关系，它可以按照一定的规律进行分布，在后面我们将看到，这些分音叠加在一起可以形成一个非周期性的波形。对于方波而言，第二个正弦分量是三次谐波，第三个正弦分量是五次谐波。其他的波形具有不同的频域特性，因为它们是由不同幅度和频率的正弦波组成的。关于波形时域和频域的详细内容请参看第 3 章。

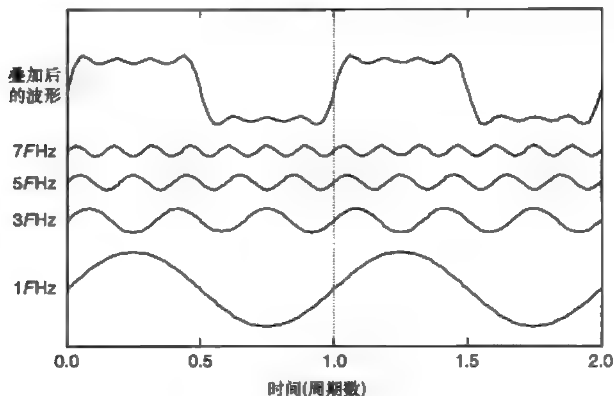


图 1.46　几个谐波关系的正弦波叠加后的波形

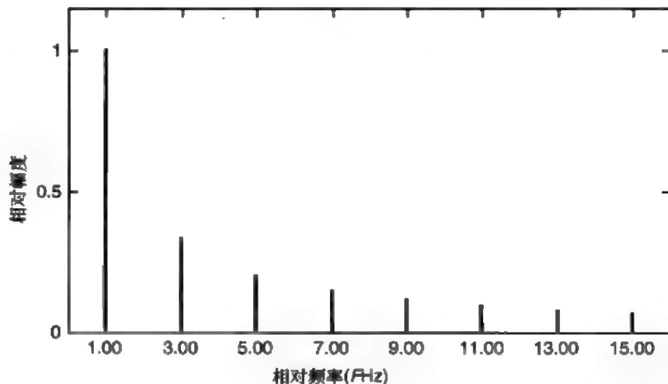
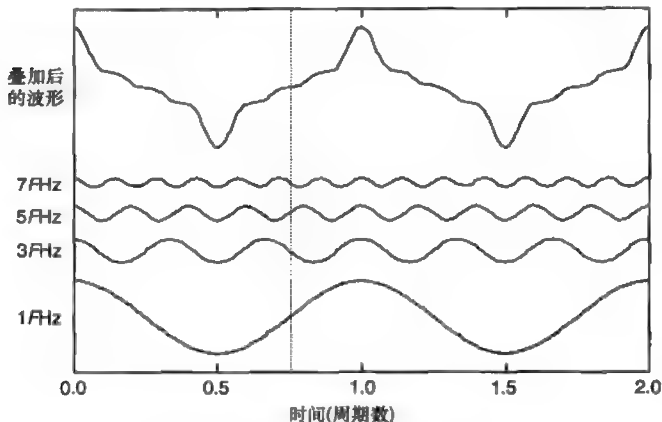


图 1.47　方波的频域表示或频谱

1.6.2 相位的影响

相位是描述每个正弦波分量初始状态的参数，它也会影响叠加后的波形。如果正弦波分量不是相加而是两两相减，即每隔一个分量的相位改变 180° ，我们可以得到图 1.48 所示的方波波形。这就是说，每隔一个分量波形平移半个波长，如图 1.48 中虚线所示。尽管时域波形是完全不同的，但是频域特性没有变化，因为频率和幅度是相同的，只有部分谐波的相位发生了改变。很多声波的波形不同，却具有相同的听觉效果。这是因为人耳听觉对声音的相对幅度比相位更加敏感，在后面章节会详细说明。但是，当相位改变足够大时，人耳就能分辨出来（请查阅 Schroeder 1975）。因此，频谱上通常只包含频率分量的幅度信息。根据人耳听觉特性，我们通常用分贝衡量幅度大小，如第 4 章中的图 4.24 就是用分贝来表示单簧管和萨克斯管演奏中央 C 的频谱。

图 1.48 改变谐波相位后的叠加效果



1.6.3 非周期性声波的频谱特性

到目前为止只讨论了周期性声波的频谱，也就是具有音高感的声波的频谱特点。然而有一些乐器，尤其是打击乐器产生的是没有音高感的非周期性信号。如何基于周期性正弦波分析这些乐器的特性呢？图 1.49 给出了问题的答案。在前面提到的合成方波的 4 个正弦波的基础上，再加入 4 个正弦波，而这 4 个正弦波选为方波谐波之间的频率，因此这些正

弦波和方波的周期性无关，但它们和其他谐波具有相同的初始相位。因为它们的频率与方波不相关，因此加入这几个正弦波分量可以破坏原有方波的周期性。通过叠加更多这样的频率分量，合成波的周期性变得越来越弱，极端情况是两个谐波分量之间充满了适当幅度和相位的非谐波分量，使合成波完全丧失了周期性。这些额外的频率分量只是在波形的开始阶段互相加强，在随后的周期将因为频率不同而互相干涉抵消。所以，在这种情况下，叠加后的声波只存在一个方波。周期性和非周期性声波频谱的主要区别在于，周期性声波具有离散的频率分量，在频谱中表现为具有一定间隔依次排列的线段形式，该间隔与波形周期成反比；而非周期性声波具有连续的频谱，不是由离散频率分量组成。但是，对于同样波形的周期性和非周期性声波，其分量幅度的包络即频谱包络的形状是相同的，如图 1.50 所示。第 3 章中的图 3.6 为小军鼓鼓刷音的非周期性波形和频谱。

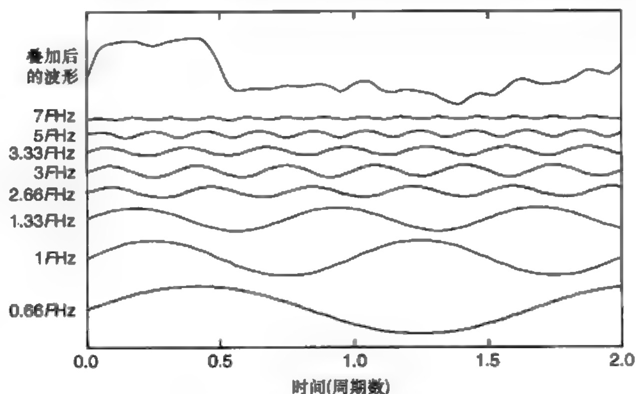


图 1.49 几个非谐波关系的正弦波叠加后的效果

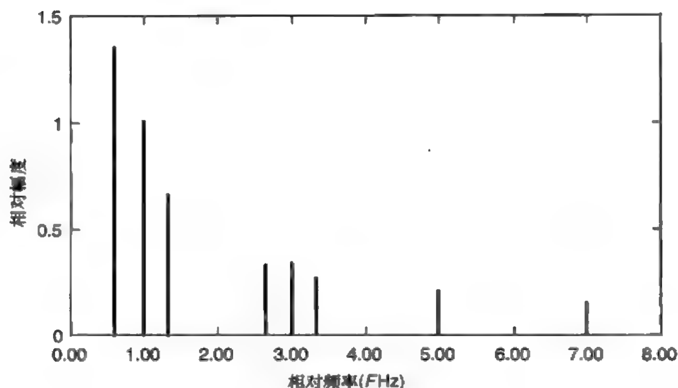


图 1.50 非周期性方波的频域表示或频谱

1.7　频谱分析

由于频谱是我们了解声音的重要途径，因此我们需要分析真实声音信号的频谱。将声音信号输入一组滤波器即可分析声音信号的频率成分，如图 1.51 所示。

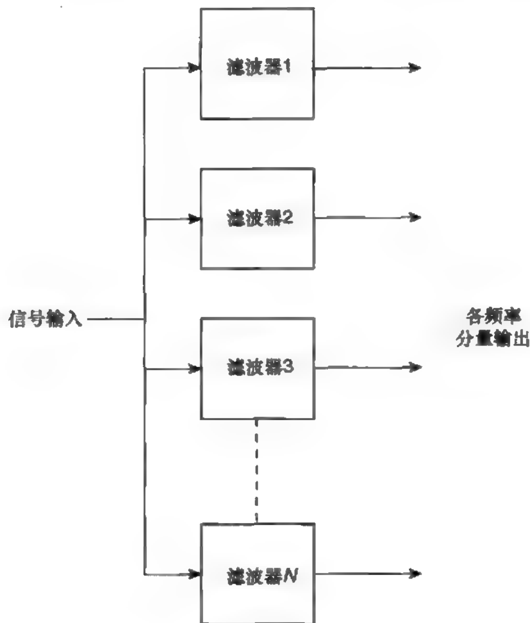


图 1.51　用于频谱分析的滤波器组

1.7.1　滤波器和滤波器种类

滤波器是从完整的声音信号频谱中分离出部分频率成分的声音信号处理器，如图 1.52 所示。滤波器主要有 4 种基本类型，是按照其对不同频率成分的作用不同来分类的，这就是所谓的滤波器频率响应。这 4 种基本类型分别为：

（1）低通滤波器：低通滤波器只能通过截止频率以下的信号；

（2）高通滤波器：高通滤波器只能通过截止频率以上的信号；

（3）带通滤波器：带通滤波器只允许通过特定频带的信号。这个特

定频带是上限频率和下限频率之间的频带，称为滤波器带宽；

（4）带阻滤波器：带阻滤波器只能通过上限频率以上和下限频率以下的信号。

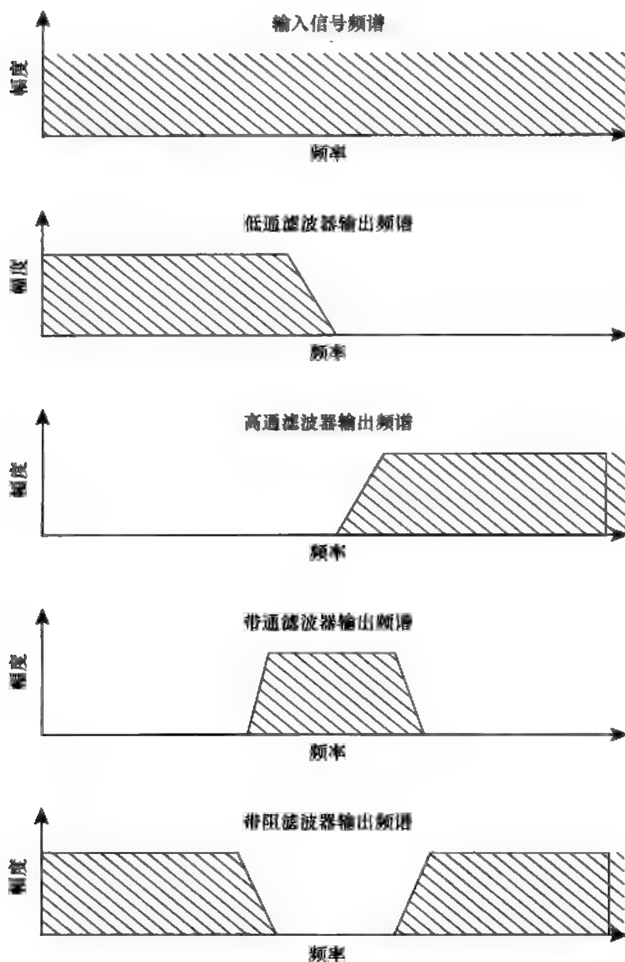


图1.52　不同类型滤波器的滤波作用

这4种类型滤波器的滤波作用如图1.52所示。通过观察可知，带通滤波器最适合于分析声波的频谱。请注意，虽然真实的滤波器会有指定的截止频率，但截止频率可以根据需要来设计，而且随着信号频率的变化，信号不会立刻被切除，而是在有限的频率范围内有一个逐步衰减的过程。

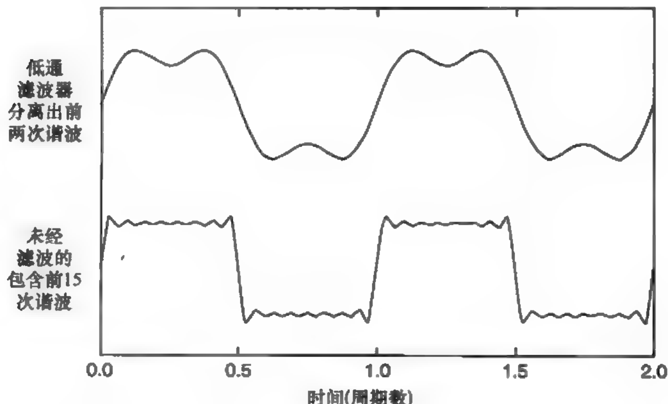
信号通过滤波器后的频谱是一个乘积的结果，因为输出信号的频谱

是滤波器的频率响应与输入信号频谱的乘积。因此如果滤波器的频率响应已知，即可利用其分析输出信号的频谱。这个技术可用于分析乐器的声学特性，如将乐器的某些特性用滤波器表示出来，这些将在后文中介绍。事实上，实现滤波的途径是多种多样的，可以是机械方法、声学方法和电学方法，很多乐器都是以声学方法对其产生的声源进行某种形式的滤波处理（见第 4 章）。

1.7.2 滤波器的时间响应

当用滤波器对信号进行频谱处理时，会产生一个问题，即滤波器改变了信号的时间特性。由于滤波器只能允许特定频段的信号通过，因此大多数滤波器存在一个时间响应。如果能够通过所有频段的信号，就不是滤波器了！通过了解傅里叶分析方法如何分析声音信号就可以了解到滤波器产生时间响应的原因。例如，如果一个声音信号通过一个低通滤波器，其上限频率是 F_{max}，那么滤波器将不会输出频率高于 F_{max} 的信号。正弦波有上升坡度这个参数，上升坡度是随频率变化的。低通滤波器输出波形的最大上升坡度是由 F_{max} 决定的，因为所有坡度大于这个上限频率的正弦波将不被输出。图 1.53 所示为方波通过低通滤波器后的波形，这个低通滤波器只允许通过前两个正弦波分量，如图 1.54 所示。[注 1]

图 1.53 低通滤波器对方波波形的影响



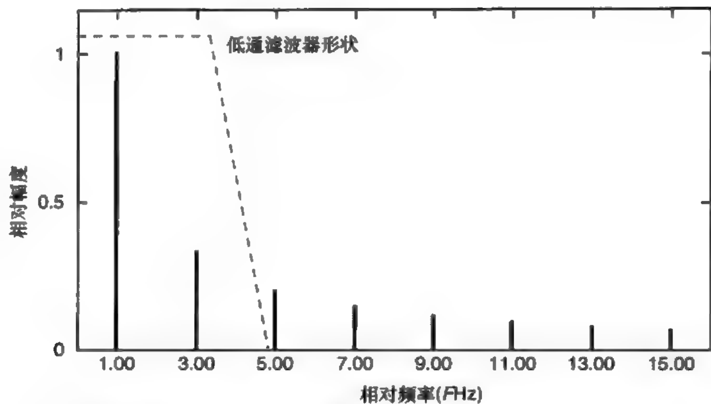


图 1.54 低通滤波器对方波频谱的影响

对于带通滤波器也可以用类似的方法进行分析。在这种情况下，滤波器给定了一个最大的频率范围，只有这个频率范围内的信号能够通过。虽然通带内的正弦波可以被滤波器输出，而且它们的频率可能相当高，但是它们的幅度包络不会变化很快，如图 1.55（上图）所示。这是由于幅度包络的变化速度取决于滤波器输出所包含的正弦波数量、所覆盖的频率范围，参看图 1.56 和图 1.55（下图）（译者注：下图的包络变化比上图快）。带通滤波器输出的幅度包络变化不仅携带着这些关于频谱的信息，而且也携带着滤波器固有的时间响应特性，它是滤波器带宽的函数。（译者注：图 1.55 和图 1.56 所指的方波实际上是被正弦波调制后的方波。设方波信号为 $f(t)$，则调制后的信号为 $\sin\omega t \cdot f(t)$，才能得到图 1.56 所示的频谱。）

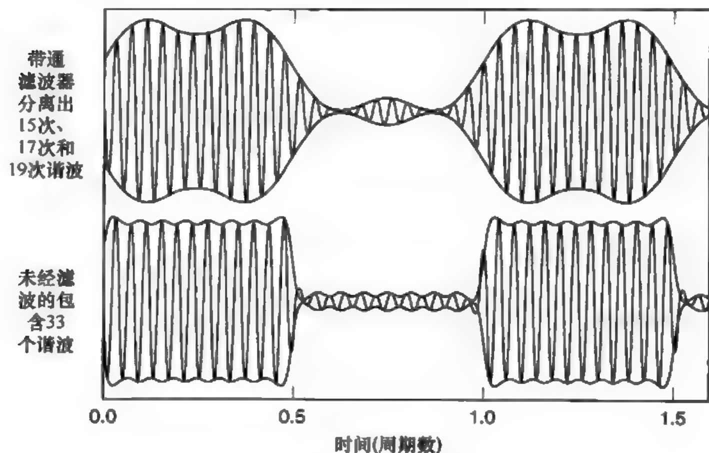


图 1.55 带通滤波器对方波包络的影响

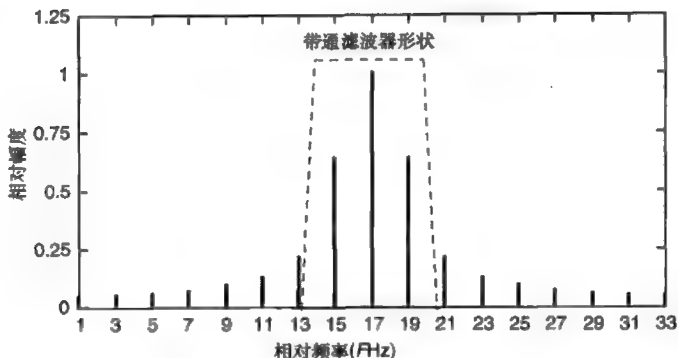


图 1.56　带通滤波器对方波频谱的作用

因此，所有滤波器都存在一定的时间响应，它是带宽或所通过信号的频率范围的函数。要注意的是，时间响应是滤波器固有的问题，不能通过技术手段加以解决。滤波器的时间响应与带宽成反比，也就是说，窄带滤波器具有较长的上升时间和下降时间，而宽带滤波器具有较短的上升时间和下降时间。在实际应用中，如果频谱分析的频率分辨率比较高，则滤波器带宽较窄并且时间响应比较差，有明显的拖尾效应。反之，如果滤波器的时间响应较好，则频率分辨率要降低，因为这时滤波器具有较大的带宽。

[注 1] 还有一种特殊类型的滤波器称为“全通”滤波器，它绝不是简单的一根导线！它可以在不改变信号频率分量幅度的情况下改变信号频率分量的相位，而在改变相位的同时也改变了其时间响应。

1.7.3　声学系统的时间响应

上述观点可以反过来这样理解，当声学系统如乐器的输出变化很慢时，其频谱覆盖的带宽较窄；而当其输出变化很快时，其频谱覆盖的带宽较宽。当波形变得充满尖峰时，很可能是由于频谱中高次谐波数量的增加。图 1.57 所示为增加方波的谐波分量后的波形变化。正如前面所述，这是由于增加了包含声音谐波分量的带宽，使得声音包络的变化更加迅速。图 1.58 和图 1.59 展示了这种效果，图中对两个相似的系统进行比较，一个幅度衰减缓慢而另一个幅度衰减较快，后者是由于系统的能量损失更大。由图可以清楚看出，衰减更快的系统允许通过更多谐波、工作频带更宽，从而使声音信号包络的变化更快。

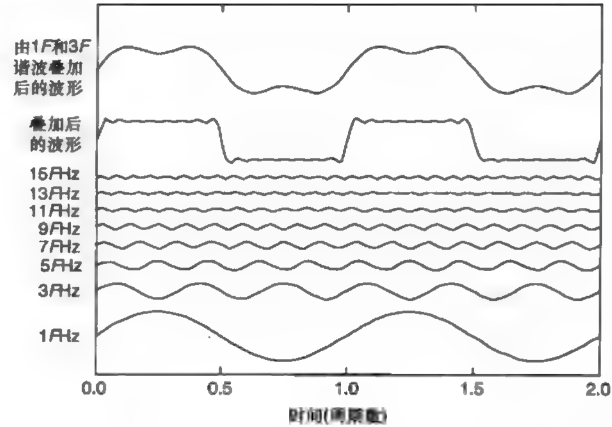


图 1.57　增加谐波分量数目对方波波形的影响

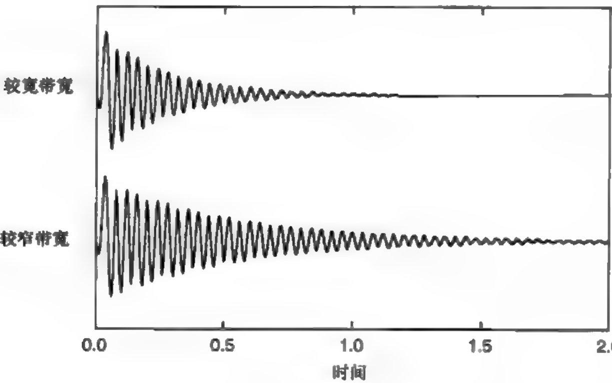


图 1.58　两个不同带宽系统的衰减率

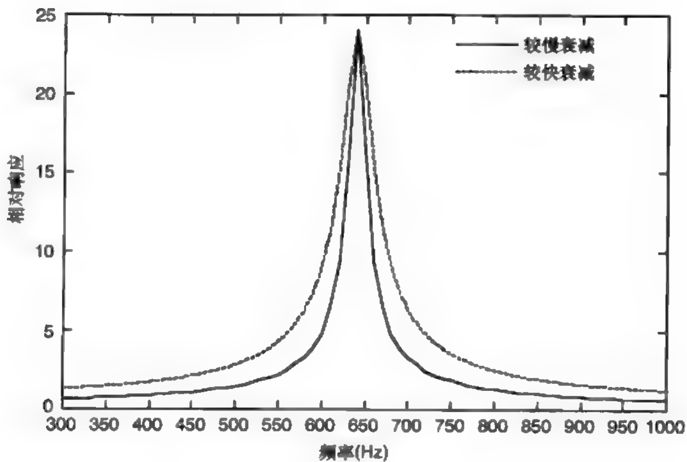


图 1.59　不同衰减时间的两个系统的频率响应

1.7.4　声音的时域和频域表示法

图 1.60 所示为一种将声音信号的时间特性和频率特性同时表示出来的常用方法，这种表示方式称为声谱图（spectrogram）。在声谱图中，信号的频谱是时间的函数，坐标轴的横、纵坐标分别表示时间和频率。信号的幅度大小用灰度表示，其中深色表示较高幅度，浅色表示较低幅度。

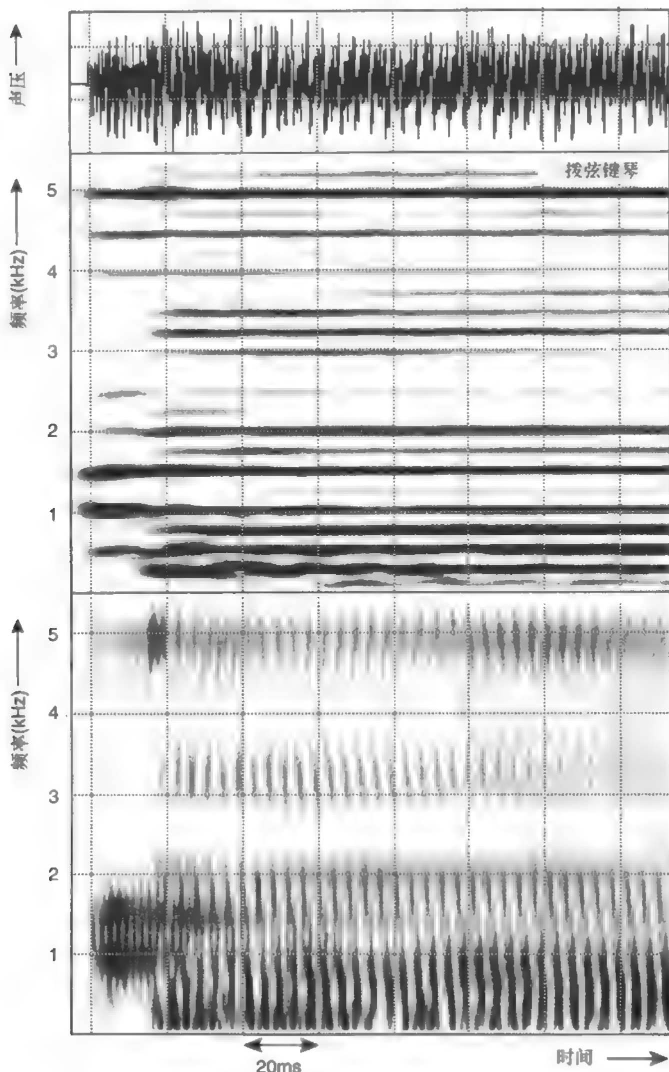


图 1.60　拨弦键琴演奏中央 C 的声压波形以及窄带（40Hz 分析滤波器带宽）声谱图（上图）和宽带（300Hz 分析滤波器带宽）声谱图（下图）

这种表示方法非常便于观察乐器的频谱随时间变化的情况，但也存在前面提及的滤波器时间响应的影响问题。图 1.60 同时画出了拨弦键琴演奏中央 C 的开始阶段分别经过窄带和宽带滤波器分析后的输出情况。在窄带滤波中，虽然谐波成分清晰可见，但是信号的初始阶段有明显的灰度由浅变深的过程，这是因为窄带滤波器具有较慢的时间响应。相反地，宽带滤波具有极好的时间响应，甚至能够表现出音高所固有的幅度的周期性变化规律，这在声谱图中表现为纵向条纹，但是却不具备分解信号谐波成分的能力。

理想的滤波器同时要满足时间响应快和带宽窄这两个特点，在物理上是不可能实现的。但是，正如我们将要看到的，人类听觉系统可以兼顾两者，使听觉既具有较高的频率分辨力又具有较好的时间响应。

本章我们探讨了有关声波波形及其特性的许多问题。然而，声音只有被双耳听到才会有意义，因此，为了全面了解乐器的性质以及音乐信号的处理技术，我们必须了解双耳的听觉机理。这是下一章将要介绍的内容。

参考文献

Beranek, L.L. (1986). Acoustics. New York: Acoustical Society of America.

Everest, F.A. (1994). The Master Handbook of Acoustics. New York: Tab Books.

Kuttruff, H. (2000). Room Acoustics. London: E&FN Spon.

Schroeder, M.R. (1975) Models of Hearing Proc IEEE, Vol 63, pp 1332–1350.

第 2 章　听觉概述

心理声学是一门研究人类如何感知声音的学科。研究心理声学首先要了解人类听觉的基本结构，从而进一步了解以下几点：

（1）正常听觉系统对传输到双耳声音的影响；

（2）心理声学基本规律的发现与音乐感知密切相关；

（3）响度过大的声音是如何损坏听力的；

（4）有听力障碍的人所面临的一些听音问题。

本章将介绍人类听觉系统的主要解剖结构图。从解剖图可以看到，沿着听觉系统的传输路径，音乐信号如何到达双耳的神经纤维，再由神经纤维传输到大脑。本章还将介绍“临界频带（critical bands）”的概念，这个概念是从音调、响度和音色方面理解音乐和其他声音感知的唯一最重要的心理声学原理。

应当指出的是，许多心理声学现象是经过实验观察得到的，是在严格控制重放声源声学特性的基础上，对多个听音者听音结果进行收集得到的。这些结果往往是通过比较两种声音而得出，例如，哪一个响度较大、哪一个音高较高、哪一个更加明亮等。很多这类实验结果至今还不能用解剖学或物理学方法解释。心理声学的研究正处在发展阶段。虽然如此，这些实验结果却为了解人耳对音乐感知的本性奠定了基础，同时，对人

耳能够感知的声音最小变化的认识也为解决声音合成的一些微妙问题提供了方向。

2.1 听觉系统的构造

人耳听觉系统的构造如图 2.1 所示，它可分为 3 大部分：外耳、中耳和内耳。这 3 个部分的解剖结构及其对入射声音信号的影响将在下文介绍。

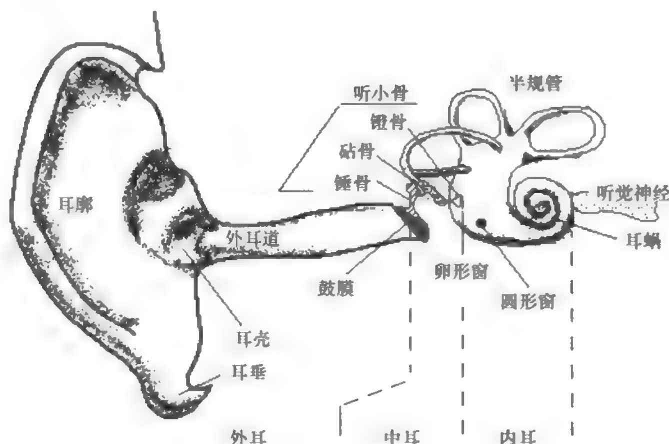


图 2.1 人耳听觉器官的主要构造

2.1.1 外耳的功能

如图 2.1 所示，外耳包括由外部皮肤组织形成的耳廓，它的形状凸凹不平。连通外耳道的入口处的凹陷结构称为耳壳。外耳道从耳壳到“鼓膜”（常称为耳鼓）的长度为 25mm ~ 35mm。外耳对进入听觉的声音的声学作用是帮助人耳对声源进行定位，同时对一些频率的声音有增强作用。

声音的定位主要是通过耳廓和耳壳完成的，耳壳还起着声学共鸣腔的作用。耳廓和耳壳的共同声学作用对于判断声源来自前方或后方以及简单判断声源来自上方还是下方是非常有效的。

外耳作为一个整体，由于受到共振的作用，会改变入射声波的频率响应。其中以外耳道的共振最为突出，其共振频率约为 4kHz。

鼓膜是一种轻、薄、高弹性的结构，它是外耳和中耳的分界面。鼓膜的解剖结构有 3 层，其中上皮层与外耳道皮肤相连；内层与中耳黏液外膜相连；上皮层和内层之间是中层，其纤维结构给鼓膜以强度和弹性。鼓膜的作用是将外界的声压变化转换成中耳的机械振动形式。

2.1.2　中耳的功能

鼓膜的机械运动通过锤骨、砧骨和镫骨三个听小骨传输到耳蜗的卵形窗，如图 2.1 所示。卵形窗将中耳和内耳连接起来。

锤骨固定在鼓膜的中层纤维上，当鼓膜处于静止状态时，这种结构使鼓膜被锤骨向内拉紧。因此，从外向里观察鼓膜呈现圆锥形凹陷形状。镫骨一侧的镫骨底板附着在耳蜗的卵形窗上。锤骨和砧骨连接得非常牢固，以便在一般强度作用下表现为一个整体。当鼓膜的振动通过球窝关节带动镫骨底板以活塞方式运动时，它们将作为一个整体转动。声学振动通过鼓膜和听小骨的机械运动传输到内耳的耳蜗。

中耳具有双重功能：（1）在不损失太多能量的情况下，将鼓膜的振动传输给耳蜗的淋巴液；（2）在外部声源或个人听力原因引起的声音响度过大的情况下，可以在一定程度保护听觉系统。

为了使能量高效地从鼓膜传输到卵形窗，通过中耳的机械振动传输方式，可以使卵形窗上的有效声压比鼓膜上的大。这是为了克服耳蜗液体较高的运动阻力（与输入端的空气阻力相比）。运动阻力可以看成是运动的“阻抗”，流体的声阻抗比空气的声阻抗大。听小骨正如一个机械的“阻抗转换器”或“阻抗变换器”，它主要通过以下两种方式实现阻抗转换：

（1）锤骨和砧骨之间形成的杠杆效应；

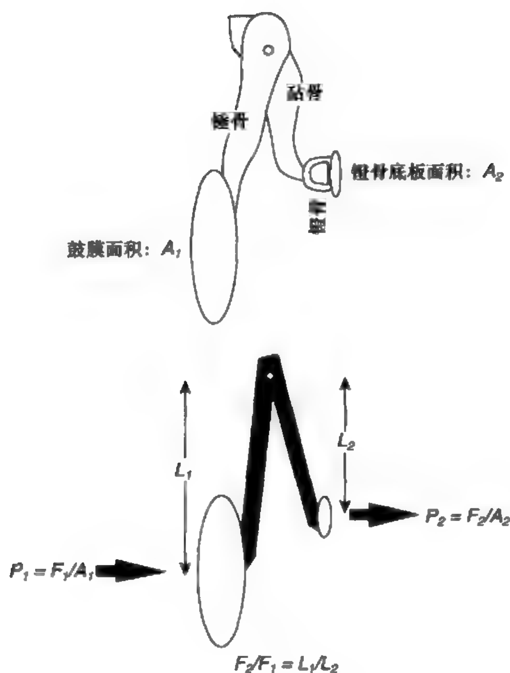
（2）鼓膜和镫骨底板面积的不同。

锤骨和砧骨产生的杠杆效应是由杠杆的长度决定的。如图 2.2 所示，镫骨底板与鼓膜的压力比与锤骨和砧骨的长度比满足以下公式：

$$F_1 \times L_1 = F_2 \times L_2$$

其中，F_1 为鼓膜的压力，F_2 为镫骨底板的压力，L_1 为锤骨的长度，L_2 为砧骨的长度。

图 2.2　中耳中听小骨的功能



因此：

$$F_2 = F_1 \times \frac{L_1}{L_2} \tag{2.1}$$

鼓膜和镫骨底板面积的差异直接影响了由鼓膜上的入射声压产生的作用在镫骨底板的压强大小。压强是指单位面积的压力，可表示为：

$$压强 = \frac{压力}{面积} \tag{2.2}$$

人耳的鼓膜和镫骨底板的面积在图 2.2 中分别用 A_1 和 A_2 表示。

鼓膜的压强（P_1）和镫骨底板的压强（P_2）可表示为：

$$P_1 = \frac{F_1}{A_1}$$

$$P_2 = \frac{F_2}{A_2}$$

因此，压力可以用压强表示为：

$$F_1 = (P_1 \times A_1) \tag{2.3}$$

$$F_2=(P_2 \times A_2) \tag{2.4}$$

将式（2.3）和式（2.4）代入式（2.1）得：

$$\frac{P_2}{P_1}=\frac{A_1 \times L_1}{A_2 \times L_2} \tag{2.5}$$

Pickles（1982）指出了与阻抗转换有关的中耳的第三个作用，也就是鼓膜本身的屈曲运动使得锤骨的受力增加到鼓膜处的 2 倍。

人耳鼓膜的面积（A_1）大约是镫骨底板面积（A_2）的 13 倍，镫骨的长度大约是砧骨长度的 1.3 倍，鼓膜的屈曲运动有一个压力的增益系数 2。因此，镫骨底板压强（P_2）大约为鼓膜压强（P_1）的 33.8（$13 \times 1.3 \times 2=33.8$）倍。

例 2.1 试用分贝表示镫骨底板和鼓膜的压强比。

压强比为 33.8 ：1，用式（1.12）可将压强比转换为分贝的表示形式：

$$\mathrm{dB}(SPL)=20\lg\frac{P_2}{P_1}$$

将压强比 33.8 代入上式得：

$$20\lg 3.8=30.6\mathrm{dB}$$

中耳的第二个功能是，在外部声源或个人听力原因引起的声音响度过大的情况下，保护听觉系统不受损伤。这与中耳的两块肌肉运动有关，分别是鼓膜张肌和镫骨肌。在声压级大约为 75dB 时，这两块肌肉会自动发生响应，并且通过听骨链变硬的过程增加中耳的阻抗。这样的生理反应可以降低振动从鼓膜传输到内耳的传输效率，从而在一定程度上保护内耳不受损伤。上述的保护机理大约可衰减 12dB ~ 14dB 的声压级，但是它只对低于 1kHz 的低频信号有效。这两块肌肉是根据它们连接的听骨链的部位而命名的。例如，鼓膜张肌是依附在鼓膜上，镫骨肌位于镫骨处。

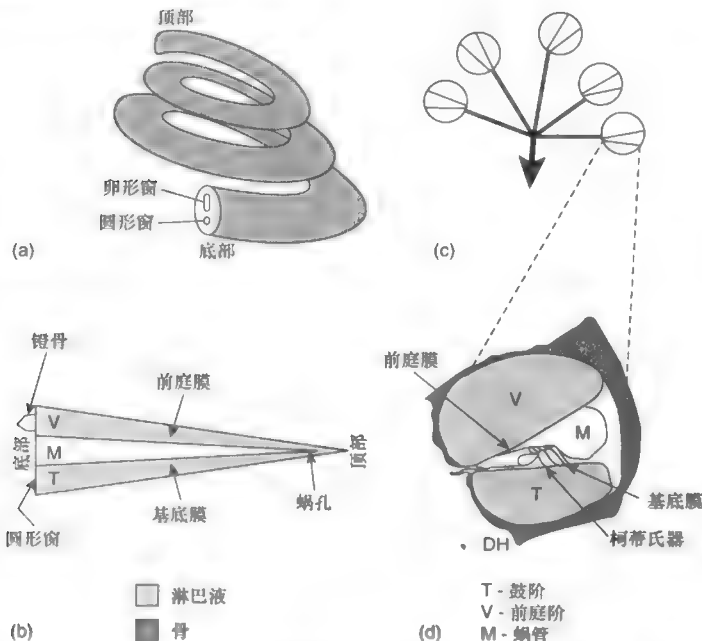
这种效应称为“听反射（acoustic reflex）”。肌肉需要 60ms ~ 120ms 的反应时间对响度较大的声音产生响应。对于响度大的脉冲声信号，如枪声，有一种说法是听反射反应时间太慢，不足以保护听觉系统。而实际上在射击的情况下，在射击声出现之前的 120ms，声音的响度已足以触发听反射，同时这样的响度又不至于损坏听觉系统。

2.1.3 内耳的功能

内耳的主要组成部分是耳蜗。耳蜗的外形有点像蜗牛壳，因此而得名。耳蜗的功能是将机械振动转换为最终传输给大脑处理的神经电脉冲，机械振动是从中耳镫骨底板通过卵形窗传输到耳蜗的。

耳蜗由一根骨质小管盘旋约 2.75 圈形成，如图 2.3(a) 所示。靠近卵形窗和圆形窗的一端是“底部”，另外一端为“顶部”，如图 2.3(b) 所示。图 2.3(c) 是耳蜗螺旋结构垂直截面图。通过 (d) 部分可以看出这个软管状结构内部被前庭膜和基底膜分为 3 个部分，其中前庭阶(V)和鼓阶(T)位于外侧，内部充满了淋巴液；内侧是蜗管 (M)，前庭阶的末端是卵形窗，骨阶的末端是圆形窗。一个理想的耳蜗的展开形状如图 2.3(b) 所示，在耳蜗的“顶部”有一个小孔，叫做蜗孔，前庭阶与鼓阶的淋巴液在蜗孔处相通。

图 2.3 (a) 耳蜗的螺旋结构；(b) 耳蜗的展开图；(c) 耳蜗的垂直截面图；(d) 耳蜗管内部结构



输入的声学振动引发镫骨底板在卵形窗产生活塞式振动，活塞式振动将带动耳蜗内淋巴液的流动。由于淋巴液基本上不可能被压缩，为补偿卵

形窗的运动，圆形窗的外膜也会运动。镫骨底板在卵形窗的向内运动将会引起圆形窗向外运动；相反，镫骨底板在卵形窗的向外运动将会引起圆形窗向内运动。这些运动会在前庭阶产生行波，并转移到前庭膜和基底膜。

基底膜的作用是对输入声音信号进行频谱分析。它的形状在耳蜗底部狭窄而薄，沿着“顶部”方向逐渐变宽、增厚，如图 2.4 所示。图 2.4 的上面部分是理想的基底膜形状，它沿着展开的耳蜗分布（可对照图 2.3（b）），可以看出基底膜在耳蜗“底部”是最窄而薄的一端，沿着“顶部”方向逐渐变宽变厚。基底膜在受到声音信号刺激后会发生响应，产生振动。由于小尺寸物体对高频信号的响应比大尺寸物体要好（试比较小提琴和低音提琴，或比较钢琴的高音弦和低音弦），所以基底膜的窄而薄的部位（底部）对高频响应最好，基底膜的宽而厚的部位（顶部）对低频响应最好。基底膜的宽度和厚度沿着耳蜗逐渐变化，输入的不同频率的纯音信号将在基底膜的不同位置激发一个最大振幅。图 2.5 用部分基底膜长度说明了这个现象，它是听觉系统对声音的“部位（place）”分析理论的基础。图 2.5 表示基底膜对 5 个不同频率的理想纯音信号的振幅包络。如果输入的声音信号是一个复合频率信号，那么整个基底膜的响应就是每一个分量的响应之和。基底膜在耳蜗底部受到刺激（见图 2.3），而底部对高频响应最好。要注意的是，对于纯音信号（或复音信号的一个分量），其振幅包络不是对称的（见图 2.5），基底膜对高频信号的恢复速度比低频信号慢，这个问题还将在第 5 章涉及。

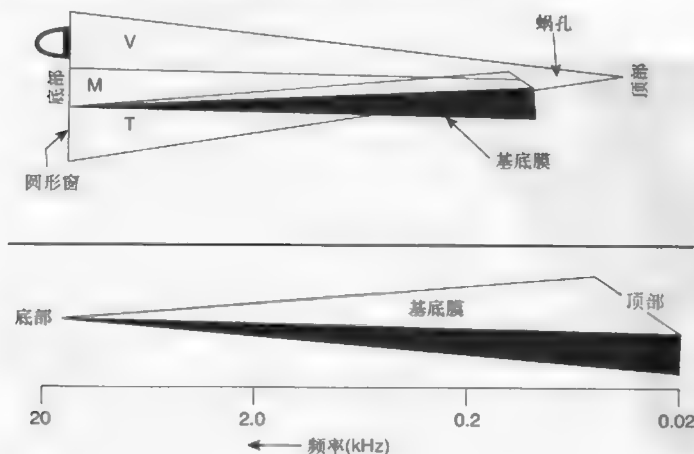


图 2.4　位于展开耳蜗内的理想基底膜形状（上图）和基底膜对不同频率的响应（下图）

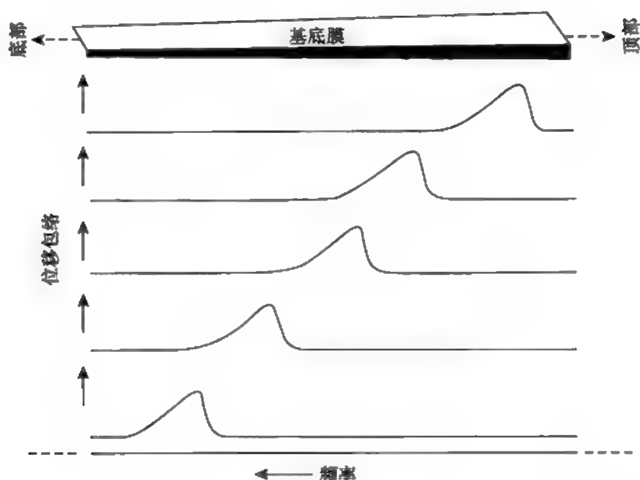


图 2.5　基底膜对 5 个不同频率的纯音信号的理想振幅包络

继 Békésy（1960）首次开创性研究后，一些研究人员已经观测到基底膜对输入不同频率正弦信号的振动情况。他们已经证实，随着输入信号频率的改变，基底膜最大位移的位置也在改变。此外还证实了，基底膜顶部到发生最大位移处的直线距离与输入频率的对数成正比，因此图 2.5 中频率轴也应该是对数形式的。从该图可以看出，频率是沿着从右至左的方向提高的，低频最大位移靠近顶部，高频最大位移靠近底部，由此保持输入声学信号从左流向右的感觉，并且还可以增强对内耳解剖性质的理解。内耳中进行低频频谱分析的部位是耳蜗的顶部，其距离卵形窗最远，卷曲在耳蜗螺旋状结构的中心。

基底膜的振动将进一步转化成神经电脉冲，并传导到大脑进行处理。基底膜附着的柯氏螺旋器官上有大量的毛细胞，毛细胞分布在基底膜周围。在受到声音信号激励时，毛细胞会发生变形，毛细胞发生变形后会触发形成神经脉冲信号。毛细胞的神经形成了螺旋束结构，这就是所谓的听觉神经。听觉神经在耳蜗的形态如图 2.1 所示。

2.2　临界频带

第 2.1 节介绍了在基底膜的机械振动机理作用下，内耳对声音进行频

率分析的原理，这是听觉“部位学说（place theory）”的基础。“部位学说”下一个需要考虑的重要问题是听觉系统对输入声音信号的频率分析能够到达什么程度。这将为理解听觉系统的频率分辨率打下基础，同时也将为研究人耳感知音乐、语声和其他声音的心理声学相关问题提供必要的支持。

输入声音信号的每一个分量都将导致基底膜在某一个特定位置发生位移，正如图 2.5 所示。每个频率分量产生的位移在峰值两侧有一定程度的延展。两个具有相似振幅、频率相近的声音信号是否能被听觉分辨，取决于每个分量在基底膜产生的位移包络是否可清晰地分离。

请参考本书光盘的第一条听音项目。假设两个纯音信号或正弦波叠加在一起，振幅分别为 A_1 和 A_2，频率分别为 F_1 和 F_2。如果保持 F_1 频率不变，而从等于 F_1 的频率开始慢慢向高或向低改变 F_2 的频率，听觉一般能听到的结果如图 2.6 所示。当 F_1 和 F_2 频率相同时，只能听到一个音高；当 F_2 高于或低于 F_1 时，可以听到声音的振幅有波动性变化，称为拍音。拍音的频率等于 F_2-F_1，或者如果 F_1 大于 F_2 时，等于 F_1-F_2。拍音的振幅在 A_1+A_2 和 A_1-A_2 之间变化，如果 A_2 大于 A_1，则振幅在 A_1+A_2 和 A_2-A_1 之间变化。请注意，当振幅 A_1 和 A_2 相等时，拍音的振幅在 $2A_1$ 和 0 之间变化。

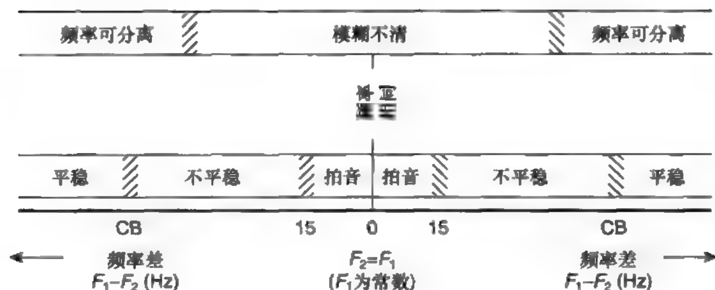


图 2.6　当纯音频率 F_1 固定而改变纯音 F_2 频率时听感的变化

当频率差小于 12.5Hz 时，大多数听音者会察觉到拍音现象。当频率差大于 15Hz 时，听感逐渐从拍音变成一个音高不明确的不稳定的声音。当频率差继续增大到达某一值时，两个频率从混合在一起变为可被人耳分离，但是仍然有不平稳的听感。当频率差进一步增大时，听感将从不平稳转变成平稳。如果两个纯音的频率差继续增大，只要在听音者听音频率范围内，则可分离而平稳的听感将持续下去。

图 2.6 中所示的听感从两个频率混合在一起到可分离、从发生拍音到不平稳再到平稳的分界线并不明确，这是因为每个听音者的听觉感知能力不是完全相同的。然而，听音者的听音能力是具有一定规律和共性的，对大多数心理声学现象都能找到共性。一般采用对大范围人群进行测量后的平均值作为最后的测评结果。

当频率差增大到某一值，基底膜上的最大位移开始变为两个峰值位移时，听觉就可以分离出两个频率的纯音信号，而不是两个信号混合在一起。然而，基底膜从一个峰值变化到两个峰值的过程将引起两个频率发生相互干扰，给人以不平稳的听感。只有当基底膜充分分离出两个峰值，使得听感从不平稳变到平稳时，才能够明确分离出两个频率。听音者开始察觉两个纯音信号从不平稳可分离转变为平稳可分离的频率差对应的带宽称为“临界频带（critical bandwidth）”，在图中标记为 CB。Scharf（1970）对临界频带给出了更正式的定义：“临界频带是主观听感突然发生明显变化的频带宽度。”

为了使临界频带这个概念更具有实用性，Glasberg 和 Moore（1990）提出了从滤波器中心频率计算等效临界带宽的公式。他们定义了具有理想矩形频率响应的滤波器，这种滤波器和听觉滤波器所通过的能量是相同的，因此称其带宽为等效矩形带宽或 ERB（译者注：ERB 是等效矩形带宽的英文缩写）。ERB 是直接计算临界带宽的一种方法，Glasberg 和 Moore 提出的适合计算任何中心频率的临界带宽公式为：

$$ERB=\{24.7\times[(4.37\times f_c)+1]\}\text{Hz} \tag{2.6}$$

其中，f_c 为滤波器中心频率（kHz），ERB 为等效矩形带宽（Hz）。上述计算公式适用的频率范围是 $100\text{Hz}<f_c<10\,000\text{Hz}$。

图 2.7 所示为根据式（2.6）描绘出来的临界带宽随中心频率变化的特性（虚线），其他直线分别代表带宽为 1 个、2 个、3 个、4 个和 5 个半音（或半音、全音、小三度、大三度和纯四度）随中心频率变化的特性，列举在同一图中便于比较。演播室通常使用 1/3 倍频程滤波器来近似代替临界频带，如图中 4 个半音带宽的直线所示（每倍频程有 12 个半音，因此 1/3 倍频程有 4 个半音）。为方便起见，键盘的琴键与滤波器中心频率轴的对应关系也在图中表示出来，小圆点表示中央 C。

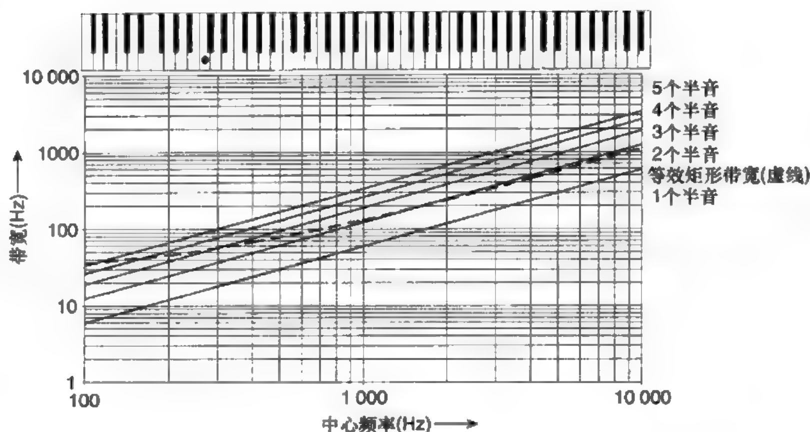


图2.7 等效矩形带宽随滤波器中心频率变化特性（直线分别表示带宽为1个、2个、3个、4个和5个半音的频率特性，中央C用小圆点标注）

例2.2 试计算中心频率为200Hz和2 000Hz的临界带宽。

将中心频率200Hz和2 000Hz分别代入式（2.6）（请注意，公式中中心频率单位为kHz，所以应该分别将0.2kHz和2kHz代入公式计算）得到，临界带宽（ERB）为：

$$200\text{Hz 时，} ERB_{200\text{Hz}}=\{24.7\times[(4.37\times0.2)+1]\}=46.3\text{Hz}$$

$$2\,000\text{Hz 时，} ERB_{2000\text{Hz}}=\{24.7\times[(4.37\times2)+1]\}=241\text{Hz}$$

如果改变图2.6中的F_1频率，则可以得到另一个CB值，由此得到临界频带随中心频率变化的特性。实际上，我们通常利用听觉的“掩蔽效应”求临界频带（见第5章），用这种方法听音者可更为清楚地感受到声音的某种“突然变化”。

单个听觉滤波器的响应特性如图2.8最下方的曲线所示，其纵轴标注为“滤波器响应”（请注意，图中横坐标从右至左频率依次增加，以便与图2.5中基底膜发生位移的坐标保持一致）。图中其他几条曲线表示输入频率间隔为f的纯音信号时理想基底膜位移包络，f的大小是垂直短线段之间的间距。滤波器的中心频率为F_c，用实线表示，同时也代表基底膜上对F_c发生响应的位置。滤波器响应曲线是根据每一个纯音信号在与F_c对应的基底膜位置所激发的位移大小描绘出来的，以此作为中心频率为F_c的听觉滤波器的频率响应。当输入纯音频率逐渐增大到F_c时，基底膜的位移沿着位移曲线中较平缓的一侧逐渐增大；当输入纯音频率继续增

大时，基底膜的位移沿着位移曲线中较陡峭的一侧迅速衰减。由图可知，滤波器响应曲线的特性是由基底膜的性质决定的，该曲线是频率 F_c 的基底膜位移曲线成镜像对称。图 2.9(a) 是按照常规（频率从左至右逐渐增大）描绘的听觉滤波器响应曲线，以便在第 5 章介绍与听觉滤波器响应不对称性有关的心理声学现象。

图 2.8 由 F_c 邻近的纯音信号产生的位移包络导出中心频率为 F_c 的听觉滤波器频率响应

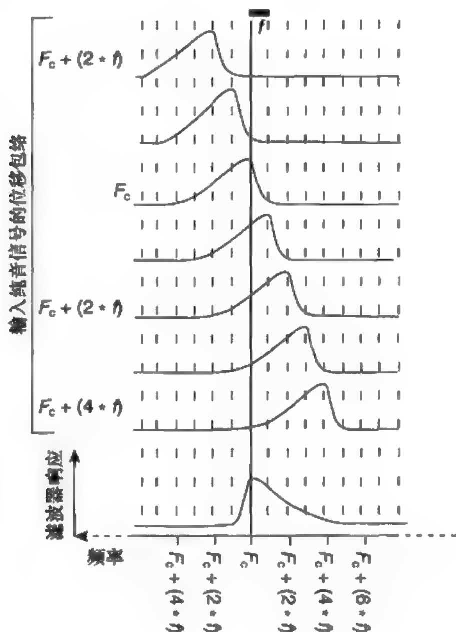
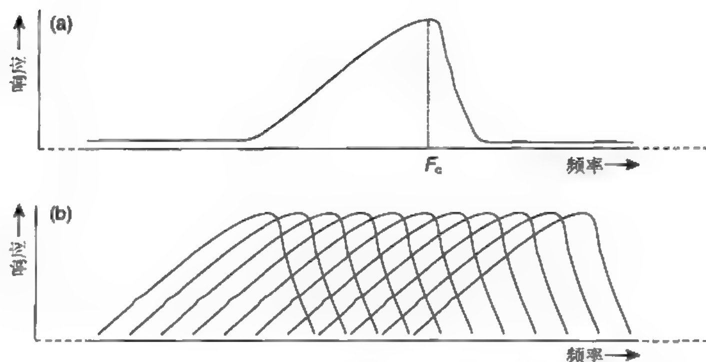


图 2.9 （a）按频率从左到右逐渐增大描绘的中心频率为 F_c 的理想听觉滤波器频率响应；（b）模拟基底膜频谱分析能力的理想带通滤波器组



基底膜的作用相当于很多频率响应交叠的带通滤波器或一个带通滤波器组，滤波器组中的每个滤波器都有一定的带宽（见第 1 章）。基于图 2.9

（a）中理想滤波器响应曲线，图 2.9（b）给出了一组听觉滤波器的响应特性。每个滤波器都具有非对称的频率响应曲线，高频一侧比低频一侧衰减得快，子滤波器的带宽就是中心频率对应的临界带宽（见图 2.7）。滤波器的重叠程度不可能做到十分明确，在实际应用中一般需要进行一定的折衷考虑。以录音室中使用的 1/3 倍频程图示均衡器为例，其子滤波器是在频响曲线 -3dB 转折频点处进行交叠的。

根据图 2.6 所示的人耳对两个纯音信号的感知能力，图中“临界频带”可以看成是滤波器组中的一个带通滤波器的带宽，其中心频率是所聆听的两个纯音的中间频率。这样的考虑忽略了基底膜位移振幅包络的不对称性（见图 2.5）以及由此产生的单个滤波器响应曲线的不对称性（见图 2.9（a）），但是这是一个很好的近似计算方法。这个滤波器（和其他中心频率与它相近的滤波器）在听觉感知到“拍频”、“不平稳混合在一起”、“不平稳可分离”时，能够将两个纯音频率包含在频带内；当听觉感知到声音从不平稳变化为平稳时，两个纯音的频率间隔变大，以致不能同时包含在这个或其他听觉滤波器的频带之内。此时，没有听觉滤波器能够同时通过这两个频率，但是，却存在听觉滤波器能够分别使两个频率单独通过。因此，听觉能够分辨出两个不同的频率，获得平稳可分离的听感。

乐音是由多个频率成分组成的。了解临界频带的听觉机理在聆听乐音时频率分析的作用，是心理声学研究的基础。听觉对声音中单个频率成分或正弦波的分辨力对于从心理声学上理解听觉如何感知音乐是非常重要的，这些主要包括旋律、和声、和弦、调音、音调、音乐动态、不同乐器的声音、声音的融合、合奏以及不同乐器同时演奏时乐声之间的相互作用等。

2.3 频率和声压级可听范围

通常人类听觉系统的可听频率范围为 20Hz ~ 20 000Hz，然而人与人之间的听觉差别是很大的。人类的可听频率范围随着年龄的增大而逐渐改变，这是人类衰老的生理表现之一，主要表现在可听上限频率的降低。

健康儿童的上限频率可达 20 000Hz，但是到 20 岁以后上限频率可能下降到 16 000Hz，20 岁以后随着年龄的增大上限频率会继续降低。这种现象通常称为“老年性耳聋”，这是人类衰老过程中的正常生理现象。随着年龄的增长，不仅听觉频率范围的上限会下降，听觉在各个频率的灵敏度也会降低，低频比高频的降低程度较小，如图 2.11 所示（参考本书光盘第 2 条）。此图也表明，这种听觉灵敏度自然降低、上限频率自然下降的现象对于男性更为明显。另外，其他环境因素也有可能引发听力损失。例如，长时间收听响度较大的音频节目（见第 2.5 节），尤其是收听从电子放大器系统重放出来的高声压级音频节目，无论其重放终端是扬声器还是耳机都有可能造成听力损失。

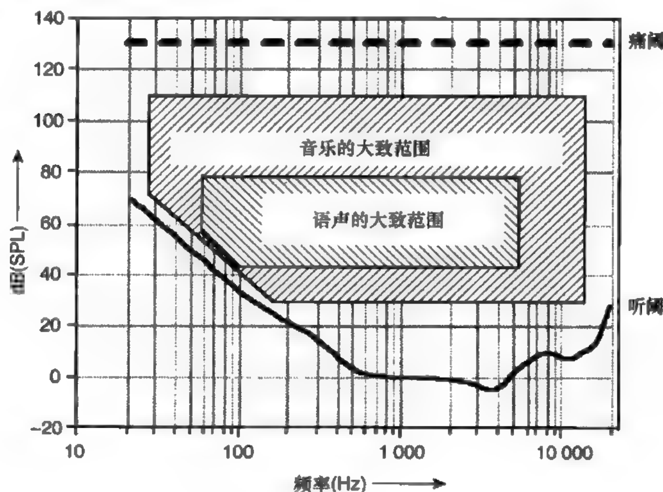


图 2.10 人耳平均听阈和痛阈曲线以及一般对话和音乐的频率和声压级范围

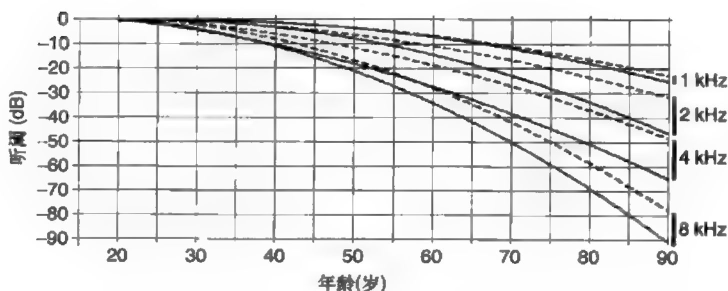


图 2.11 女性（虚线）和男性（实线）的“老年性耳聋”现象

人耳对不同声压级、不同频率声音的灵敏度在一个很大的范围内变化。平均而言，人类听觉系统可察觉的最小声压在 4kHz 附近约为 10μPa

或 10^{-5}Pa，如图 2.10 所示。不会引起疼痛感的最大平均声压约为 64Pa。响度最大的痛阈与响度最小的听阈的声压之比为：

$$\frac{\text{痛阈}}{\text{听阈}}=\frac{64}{10^{-5}}=6\,400\,000=6.4\times10^{6}$$

听觉的声压变化范围非常大，不便于处理数据，因此引入声压级（SPL）的概念（见第 1 章），其单位为 dB，换算公式如下：

$$\text{dB(SPL)}=20\lg\left(\frac{p}{p_{ref}}\right)$$

其中，p 为实际声压（Pa），p_{ref} 为参考声压（20μPa）。

例 2.3　试计算 1kHz 听阈和痛阈的声压级。

1kHz 的听阈实际上就是 p_{ref}，用分贝表示为：

$$20\lg\left(\frac{p_{ref}}{p_{ref}}\right)=20\lg\left(\frac{2\times10^{-5}}{2\times10^{-5}}\right)=20\lg1$$
$$=20\times0=0\text{dB(SPL)}$$

痛阈的声压为 64Pa，换算成声压级为：

$$20\lg\left(\frac{p}{p_{ref}}\right)=20\lg\left(\frac{64}{2\times10^{-5}}\right)=20\lg(3.2\times10^{6})$$
$$=20\times6.5=13\text{dB(SPL)}$$

由于听觉可感知的声压范围极其宽广，相差的数量级比较大，使用声压级的分贝度量方法可将数值控制在 0 ~ 130，使用起来比较方便。另一方面，用对数标度度量声波的大小更为合适，因为听觉的响度感觉更接近于对数变化规律。

听阈随着频率的变化而变化。听觉对中频的灵敏度比高频和低频都要高，图 2.10 中底部的曲线代表在 20Hz ~ 20kHz 正弦信号激励下的平均听阈。图中顶部的曲线代表痛阈的平均特性，实际上痛阈也是随着频率的变化而变化，只是变化不是很大。从图中可以看出，听阈和痛阈之间的最大声压级差（动态范围）130dB 出现在大约 4kHz 处，当频率高于或低于 4kHz 时，动态范围则明显降低。图 2.10 中还表示了语言和音乐平均统计的声压级和频率范围。表 2.1 列举了日常声音的大致声压级，以供参考。

表 2.1 日常生活中常见声源的声压级

声源	dB（SPL）	听感描述
远程射击在射击手双耳处声压级	140	
痛阈	130	糟透了！
喷气式飞机起飞（约 100m 处）	120	
夜总会舞池的最高声压级	110	
大声喊叫（1m 处）	100	非常吵闹的
重型卡车（10m 处）	90	
繁忙的汽车交通（10m 处）	80	
车内声压级	70	吵闹的
正常对话（1m 处）	60	
办公室噪声级	50	
安静地区的客厅	40	安静的
夜间的卧室	30	
空场音乐厅	20	
微风轻吹树叶	10	刚刚能听到
儿童的听阈	0	

2.4 响度感觉

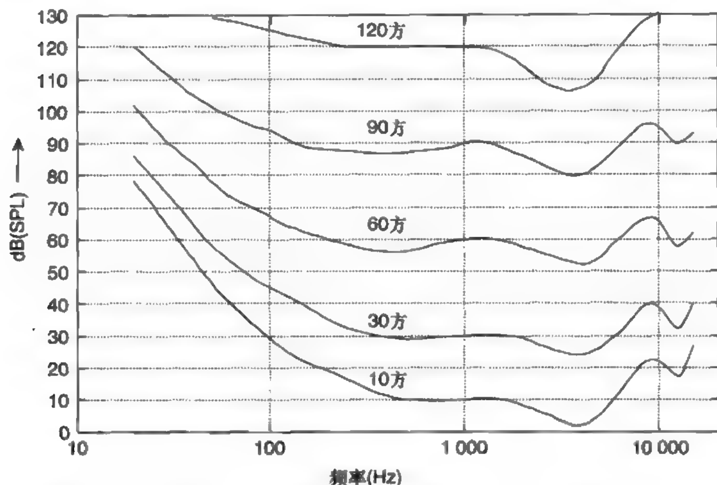
虽然声音在听觉上产生的响度感觉与振幅相关，但它们之间不是简单的一对一的函数关系。心理声学效应不仅受环境和背景的影响，而且与声音的本身性质有关。同时，心理声学效应是很难度量的，因为它与听音者对声音的描述有关。无论从道德上还是从技术上说，都不可能将一个探测器放入大脑内去测量声音的响度。

在第 1 章已经介绍了声压级和声强级，在不存在干扰现象的自由场中，声压级和声强级近似相等。听觉是一个对声压非常敏感的器官，它可以将声音信号的频谱划分成一系列带宽重叠的频带，其带宽随频率的增加而增大。然而，这是对声音的客观描述，也是对听觉功能的描述，却不能告诉我们响度感觉与客观测量的声音振幅之间的关系。对这些问

题的思考可以帮助我们理解双耳聆听音乐的某些现象。

声波的声压幅度与响度感觉并没有直接的联系，实际上声压幅度较小的声音有可能比声压幅度较大的声音具有更强的响度感觉。为什么会这样呢？因为声音的频率各不相同，听觉的灵敏度随频率的变化而变化。图 2.12 所示为人耳的等响曲线。Fletche 和 Munson(1933) 最早对等响曲线进行了测量，随后其他人也进行了测量。这些曲线表示了声压级和听觉感知的响度之间的关系，表明了当声音的声压级达到多少时，才能感觉到它与某特定声压级的 1kHz 纯音同样响。等响曲线有两个主要声学特征。

图 2.12　人耳等响曲线



第一，在大于 1kHz 时等响曲线变得凹凸不平，这是由于外耳的共振产生的。外耳道长约 25mm，一端封闭另一端开放，使得听觉在约 3.4kHz 产生第一个共振，由于外耳道形状的不均匀性，第二个共振频率约为 13kHz，如图中所示。共振现象可以提高听觉在共振频率的灵敏度。需要注意的是，灵敏度的提高是由于外耳的声学效应引起的，与信号本身的强度无关。第二，听觉的灵敏度还与声压级有关，这是由听觉内在的转换和诠释声波的方式决定的，使灵敏度频率响应随声压幅度的变化而变化。这种效应在低频尤为明显，但是在高频处也存在。以上两个现象引起的最终效果是，听觉的灵敏度是随频率和振幅变化的函数。换言之，双耳的频率响应是不平坦的，同时它也与声压大小有关。所以，两个声

压级相等的纯音信号的响度并不一定相同。例如，以一个刚刚能够听到的 20Hz 纯音信号的声压级重放 4kHz 纯音信号，响度要比 20Hz 响得多。响度相同的不同频率纯音信号具有不同的声压级，同时不同声压级声音的相对响度也与其重放的绝对声压级有关。

与频率和声压级有关的正弦波响度通常用单位“方”（“phon”）进行度量。“方”值是通过让听音者将声音与 1kHz 纯音的响度进行比较得到。一个声音的响度级（“方”值）定义为和它同样响的 1kHz 纯音的声压级。观察图 2.12 的等响曲线可以看出，*N* 方的曲线穿过 1kHz、声压级为 *N* dB（SPL）的点，声压级较大时等响曲线相对比较平坦。由于不同频率的相对响度随着声压级变化而改变，因此当重放音量变化时，声音在不同频率的平衡感也会发生变化。这个现象正如在听录音作品时，当减小音量时，低频和高频成分与中频成分相比在一定程度上被抑制，声音变得比较单薄、暗淡。最理想的情况是，在听录音作品时，用原始录音的响度来进行重放。然而，在大多数情况下这是不合常理的，特别当重放摇滚乐时，这是难以做到的，因为给摇滚乐混音的声压级常常大于 100dB！

在 20 世纪 70 年代的早期，高保真音响制造商在设备中设置了一个“响度”按钮，其功能是提升低频和高频成分，使在低音量重放时等响曲线保持平坦，以此补偿听觉在低频和高频灵敏度的降低。这种方法是错误的，主要表现在两个方面：

首先，这个方法直接利用等响曲线进行补偿，而不是利用两个不同声压级的等响曲线的差值进行补偿。后者的方法会更加准确，并且这种方法已经在专业产品中得到运用，从而可以在夜总会等场合以普通音量得到大音量重放的效果。

其次，等响曲线是用纯音信号以相近的声压级测量得到的。而真实的音乐信号包含许多振幅不同的各种频率成分，并且声压级的变化并不直接符合这些曲线的变化规律。在随后章节里，我们将介绍如何分析复音的响度。事实上，由于听觉的响应不仅和声音的绝对声压级有关，而且和频率有关，因此简单地通过提升低频和高频进行补偿是不可行的。

2.4.1 响度的测量

由于响度受声音的多方面因素影响，因此设计一个能够精确读出人耳响度感知的仪表是十分复杂的，一般只能使用仪表对声音响度进行大致测量，这种仪器称为声级计。方法是测量频率特性加权处理后的声压级，以补偿听觉灵敏度随频率变化的特性。显然，最佳的补偿方式取决于所测量的绝对声压级大小，因此针对不同的声压级需要采取不同的计权方法。图 2.13 所示为两种常用的计权方法，即"A"计权和"C"计权。"A"计权最适于测量幅度低的声音，因为它能够大致模拟人耳对低声压级的灵敏度频率特性。"C"计权模拟的是高声压级时的听觉灵敏度频率响应，因为它比"A"计权对低频信号更加灵敏。用"A"计权测得的声压级通常用 dBA 表示，"C"计权测得的声压级用 dBC 表示。尽管"A"计权更适合于低声压级的测量，对低声压级能够较好地逼近人耳的响应，但是现在已推荐使用"A"计权作为所有声压级测量的计权方法，目的是使测量结果更具有一致性和可比性。

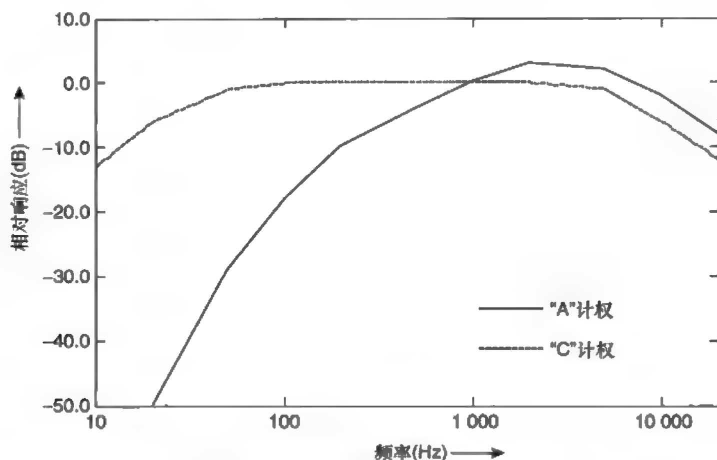


图 2.13 "A"计权和"C"计权的频率响应

使用声级计时，除了考虑频率的加权外还要考虑其他因素。要评估声压级的大小，必须对至少一个周期的波形取平均，可能的话尽可能多取几个周期。因此，声级计有"快"和"慢"的时间响应检波特性。慢的时间响应可以估计平均声压级，而快的时间响应可以跟踪声压级急剧变化的声音信号。

有时候测量一段较长时间的等效声压级是非常必要的，这对衡量人耳受到噪声的干扰程度，从而了解听觉是否由于噪声导致听力受损尤为重要。等效声压级不能通过调整声级计的“快”和“慢”时间响应来完成，需要用 L_{eq} 来测量。L_{eq} 的测量方法是将一段时间（例如 15 分钟或 8 小时）内各时刻声压的平方相加后开平方根而得出的，这可以算出一段时间的等效声压。也就是说，L_{eq} 是一个不变的声压级，它等效于一段时间内不断变化的声压级的作用。通过对它进行平方计算，L_{eq} 还是一种估计信号总能量的方法。多个短时间测量的 L_{eq} 可以叠加在一起，通过对每个 L_{eq} 的平方和开平方根后得到一个长时间的 L_{eq}，如式（2.7）所示：

$$L_{eq(total)}=\sqrt{L_{eq1}{}^{2}+L_{eq2}{}^{2}+\cdots+L_{eqn}{}^{2}} \qquad (2.7)$$

其中，L_{eq1}，L_{eq2}，…，L_{eqn} 为每个短时间 L_{eq} 的测量值。

声级计的 L_{eq} 测量功能为噪声监测提供了强有力的工具。和传统的声压级测量一样，也可以运用“A”计权和“C”计权对 L_{eq} 进行加权测量。

2.4.2 简单声音的响度

图 2.10 表示了两个响度阈限，即听阈和痛阈。正如我们已经看到的，听觉的灵敏度随着频率的变化而变化，听觉的听阈也随着频率的变化而变化，如图 2.14 所示。图中灵敏度最大值对应于声压 10μPa 或声压级 -6dB。请注意，这是听音者对正前方的声源进行单耳听音的测量结果。如果声源位于听音者头部一侧，则峰值灵敏度会增大 6dB，这是由于头部衍射使声压增大造成的。另外还有一些实验表明，双耳听音的灵敏度大约会增大 3dB ~ 6dB。在 4kHz 时，即在灵敏度峰值频率时，在常温条件下和一个临界频带范围内，由于空气分子的布朗运动（Brownian motion）引起的压强变化对应的声压级约为 -23dB。因此，听觉的灵敏度几乎达到理论上物理灵敏度的极限。换句话说，比人类听觉系统更加灵敏将没有任何意义。因为如果那样的话，我们每时每刻都能听到空气热扰动发出的嘶嘶声！很多录音室和音乐厅的设计者都在努力尝试将室内噪声控制在人耳听阈以下，这样人耳就不能察觉噪声的存在。

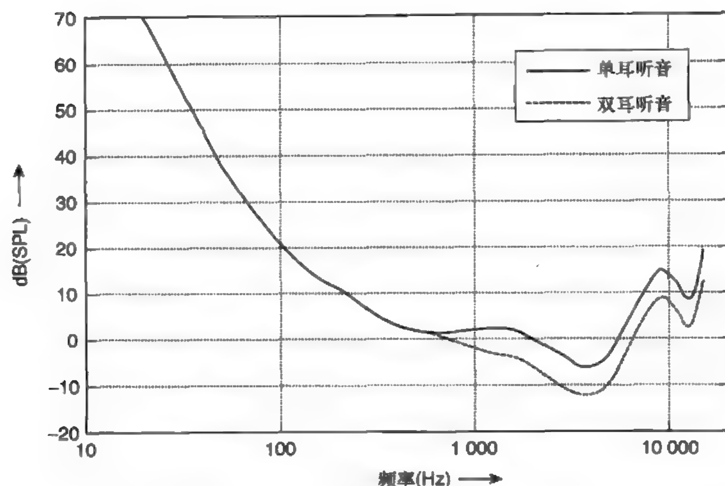


图 2.14 听阈随频率变化特性

另一个听觉阈限是幅度的最小可觉差，它与声音本身的频率和幅度有很大关系。当声压级（相对于听阈 0dB）在 20dB ~ 100dB 时，宽带噪声的幅度可觉差大约为 0.5dB ~ 1dB。低于 20dB 时，人耳对声压级的变化就不是很灵敏了。对于单频正弦波，听觉对声压级变化的灵敏度与上述截然不同，并随着幅度和频率发生明显变化。例如，1kHz、10dB 的声压级可觉差为 3dB，而 80dB 的可觉差为 0.3dB。其他频率也会发生类似的现象。总体上说，其他频率的声压级可觉差比 1kHz ~ 4kHz 的要大。上述现象使我们很难说清楚某个声音的声压级可觉差到底是多少，因为它与聆听的声音特性的细节有关。还有一些数据表明，包含一定谐波成分的声音的声压级可觉差接近于宽带的情况，大约是 0.5dB ~ 1dB，而不是纯音信号的可觉差。一般认为声音的声压级可觉差约为 1dB。

声压变化带来的声音响度的变化也与声音信号本身的性质有关。对于宽带噪声或具有多次谐波的声音，人们普遍认为声压级改变 10dB 时，响度感觉增大一倍或减半。然而，这种比例关系取决于声音的性质，并且在具体数值和有效性上还存在一定争议。

例 2.4 假设所有小提琴演奏的声压级是相同的，试计算增加多少把小提琴时弦乐组的响度可以增大为原来的两倍。

由第 1 章可知，N 个非相干声源的总声压级计算公式为：

$$p_N = p\sqrt{N}$$

用声压级表示为：

$$SPL_N = SPL + 10\lg N$$

如果响度要增加为两倍，那么声压级要增大 10dB，也就是要使 $10\lg N=10$。因此，需要弦乐组小提琴的数量增至 10 倍，响度才可以增大为原来的两倍。

响度感知除了与频率和幅度有关外，还与声音的持续时间有关。图 2.15 所示为纯音信号的响度与持续时间的关系。由图可知，当声音持续时间超过约 200ms 时，响度感觉不会发生变化；而当持续时间小于 200ms 时，响度感觉会随时间的缩短而减弱，即人耳感知的响度与纯音信号的持续时间成正比。这就是说，当我们聆听幅度变化较大的声音时，响度并不是由短暂的峰值幅度决定的，而是更多地由 200ms 内的平均声压级决定。

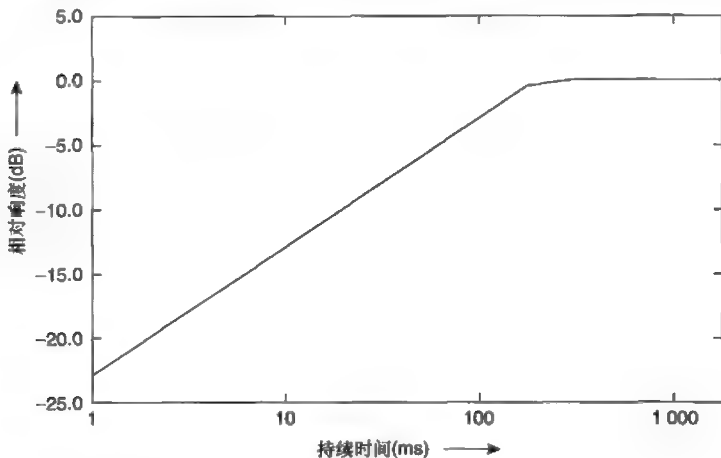


图 2.15 纯音持续时间对响度的影响

2.4.3 复杂声音的响度

真实声源一般不是纯音信号，而是包含多个频率成分。我们已经知道，人耳可以将声音信号按临界频带分离成多个频带。大脑处理临界频带以内的声音信号不同于处理临界频带以外的声音信号，因此，对于这

两种情况，响度感觉是不同的。

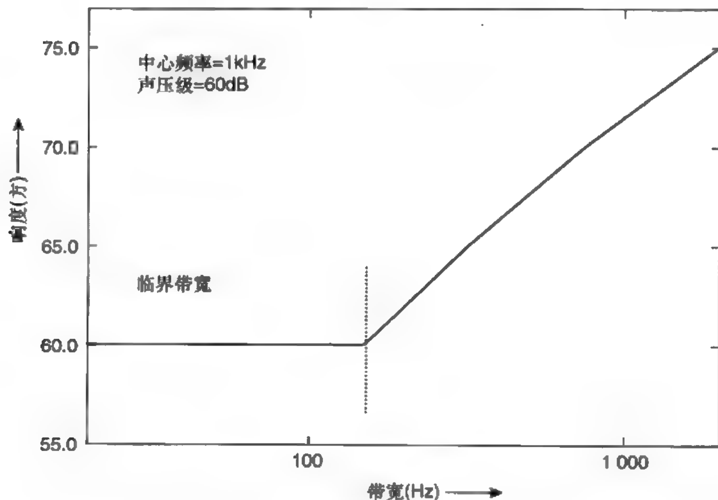
对于第一种情况，人耳可以将临界频带以内的能量进行积分，并将频带内所有声音做为一个整体进行处理。因此，当所有声能集中在临界频带以内时，响度与临界频带内的总声强成正比，即：

$$响度 \propto p_1^2 + p_2^2 + \cdots + p_n^2 \tag{2.8}$$

其中，p_1，p_2，…，p_n 分别为第 n 个频率分量的声压。

人耳对声压很敏感，而声强与声压的平方成正比（见第 1 章）。由于声强与声压的平方和成正比，在总声强不变的情况下，声音响度与临界频带内所含频率分量的数目无关。当声音的频率分量超出临界频带范围时，就会由其他临界频带的频率分量引起另外一种效应。此时，响度感知受到多个临界频带的影响，大脑将每个临界频带的响应叠加在一起。当出现这种情况时，即使总声强不变，响度感知也会随之增大。图 2.16 所示为在总声强级不变时，人耳对声源的主观响度感觉随声源频带宽度变化的特性。临界频带由耳蜗内驻波峰值的位置决定，因此临界频带内所有的能量将会在这个基底膜位置综合产生一个整体效果，形成一个相应的神经脉冲。另一方面，超出临界频带的能量将会触发其他神经产生电脉冲，这些额外的神经脉冲会引起响度感觉的增大。

图 2.16　声音带宽对响度的影响



听觉对覆盖一定频率范围的复杂声音的响度感觉还受到心理活动的

影响，使响度感知变得非常复杂。例如，听音者可能会特别关注某一个声源，如独奏、对话，此时会忽略或不那么关注其他声源，此时响度感觉倾向于由所关注的声源决定。

与纯音相似，声源的持续时间也会影响对复杂声音的响度感觉。像纯音信号一样，当复音的持续时间大于200ms时，人耳感知的响度不受声音信号持续时间的影响；当复音的持续时间小于200ms时，人耳感知的响度大小与复音信号的持续时间成正比。

2.5 噪声引起的听力损失

双耳是一个非常敏感而且准确的声音能量转换和分析器官。然而，高强度声音或噪声可能损伤听力，这种损伤主要表现在以下两个方面：

（1）听觉灵敏度的降低：噪声致使声音转换为神经脉冲的效率降低，这是由于每个柯蒂氏器中的毛细胞被损坏而造成的。请注意，这不同于为保护听觉系统而产生的听反射所引起的短时间内听阈提高的现象。如果某人处在高强度噪声环境中，这种听觉灵敏度的降低表现为听阈的提高，如图2.17所示。受到短时噪声影响时，这种听阈提高的现象可能是暂时的，但是如果受到长时间高强度声音影响时，由于毛细胞的活性降低，这种听阈提高的现象最终会变成永久性的。

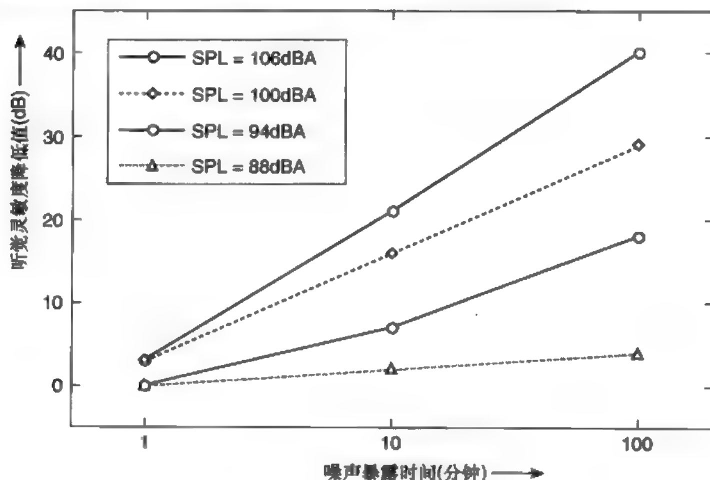


图2.17 噪声暴露对听觉灵敏度的影响（Tempest 1985）

（2）听觉分辨力受损：这是一个更加微妙的影响，但是在许多情况下比第一种更为危险。我们知道，听觉系统的听音和分析声音的能力主要表现在能够将声音分离为不同频带，即按临界频带分析声音。这些频带非常狭窄，是由于耳蜗内正反馈结构使得驻波增强而造成的。这个结构非常容易损坏，而且对噪声干扰比主传导系统更为敏感。这种损伤不仅会降低听觉滤波器的增益，还会增大听觉滤波器的带宽，图 2.18 所示为理想化情况。导致的结果主要有两种：第一，分离不同声音成分的能力减弱。由此会降低我们理解语言或从嘈杂的噪声环境中分离所需声音的能力。有趣的是，由于临界频带内可能因此存在不只一个谐波频率，乐声可能从和谐变得不那么和谐了，这个问题将在第 3 章中做进一步的解释；第二，听觉灵敏度降低，如图 2.18 所示。这是因为听觉的声音增强机理损伤也会降低双耳的灵敏度。由于这种现象很难测量和察觉，因此更为隐伏。它不仅仅表现为所察觉声音响度级的降低，更多地表现在分析和理解声音能力的降低上。

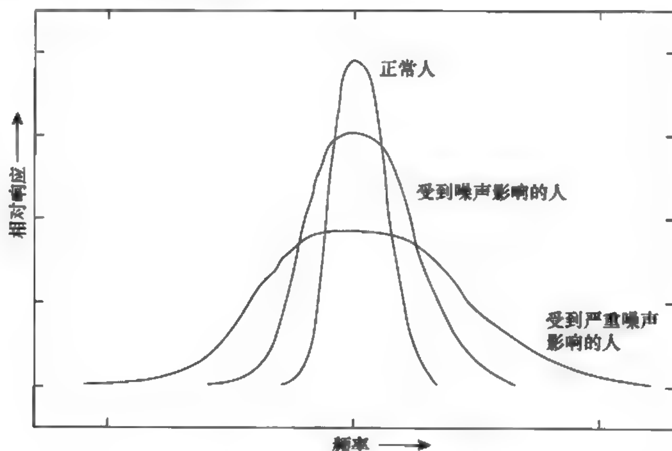


图 2.18　噪声暴露对临界频带的影响（理想化特性曲线）

另外一种由过度承受噪声引起的听力受损现象称为耳鸣。耳鸣是指耳蜗自发产生噪声的情形，该噪声可以是有音高的，也可以是随机噪声，或二者兼而有之。对于噪声性耳鸣，一旦暴露在较强的噪声中，就会引发耳鸣。同样令人烦恼的是，有数据表明患有噪声性耳鸣的人更易于患噪声性听力损伤。

由于损伤是由过度暴露在强噪声中引起的，因此较容易发生在听

觉灵敏度较高的频率上。听觉对外耳道的第一个共振频率最为敏感，大约为 4kHz，因此这也是大多数人最早出现听力损伤的频率。出现在这个频率附近的听力损伤通常表现为测听凹陷，即听力图的形状出现很明显的谷点，如图 2.19 所示。这种独特的听力测试曲线形状说明听力损失是由过度噪声暴露引起的，而不是老年性听力衰老所呈现的高频损失。

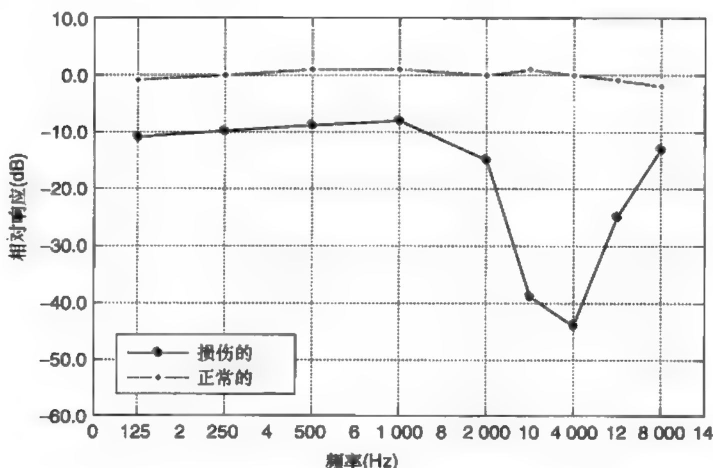


图 2.19　正常人和听力损伤者的听力图（Tempest 1985）

听觉可以承受多大的噪声呢？有数据表明，在普通嘈杂的西方社会噪声对听力造成了一定程度的永久性损伤。听力测量表明，对于特定年龄，西方人的听阈比正常听力的听阈要高一些。然而，这也可能是其他原因造成的，例如环境污染等。但是，很多数据表明，当噪声声压级大于 90dB(A) 时可能会导致永久性听力损伤。这个事实得到例如英国立法机构的认可，要求工人的作业噪声强度应低于 90dB(A)，也称为“二级行动声级”。欧洲的立法机构在 2006 年 4 月将该极限声压级下降为 85dB(A)。如果工作环境的噪声声压级大于这个数，那么雇主有义务为员工提供听力保护措施，使双耳接收到的声压级小于 85dB(A)。同时还有一个“一级行动声级”，它比“二级行动声级”低 5dB。如果工作环境声压级大于这个数字（在欧洲原来是 85dB(A)，现在是 80dB(A)），那么员工可以要求提供有效的听力保护措施。

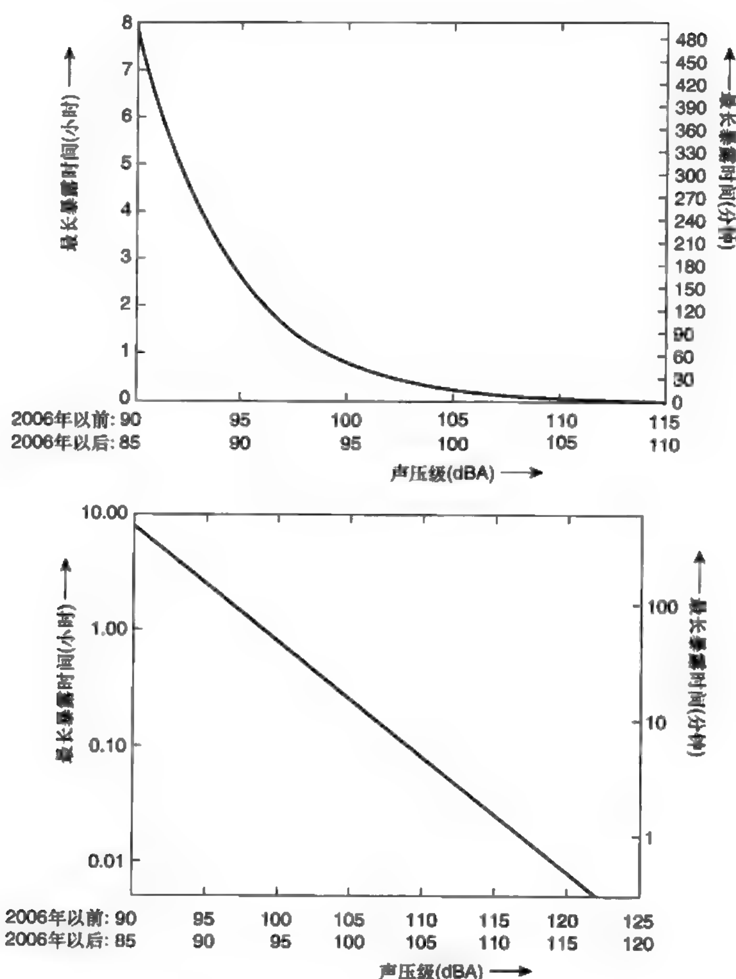
请注意，该条例因国家而异，读者应查阅当地制定的噪声限定法规。

2.5.1 累积噪声剂量

在大多数与音乐有关的场合中，短时间内噪声级可能超过 90dB(A)。例如，音乐会中听众听到的声压级很可能在瞬时高于这个值，尤其在聆听某些作品如 Elgar 的 Dreams of Gerontius 或 Orff 的 Carmina Burana 时更为明显。这种现象同样存在于很多实际工业和社会场合中，噪声级可能在短时间内大于“二级行动声级”90dB(A)(自 2006 年开始欧洲改为 85dB(A))。如何将不连续噪声和连续噪声联系起来呢？换句话说，短时间暴露在声压级为 96dB(A)的环境中对听力损伤有多大？我们可以利用评价辐射暴露相似的技术来评价噪声暴露强度，即“累积剂量”。累积噪声剂量定义为一段特定时间内的等效声压级，目前是以 8 个小时为计量时间。换言之，噪声级可以在一定时间范围内大于“二级行动声级”，只要 8 小时以内测得的等效连续声压级低于“二级行动声级”即可满足要求。评价量用前面提到的 L_{eq} 来衡量，允许 8 小时内最大等效连续声压级为 90dBL_{eq}。这意味着，听者可以在 93dB(A)的环境中暴露 4 小时，在 96dB(A)的环境中暴露 2 小时，依此类推。图 2.20 所示为 2006 年 4 月之前和之后“二级行动声级”最长暴露时间随声压级变化的线性表示和对数表示。由此可知，大于 100dB(A)的声压级只能持续很短一段时间，大约只有半小时。图中声压级具有一定上限，这是因为响度非常大的声音会可能瞬时导致鼓膜破裂，有时候可能永久性失去听觉。

这种测量噪声暴露剂量的方法没有考虑噪声的频谱，因为很难在实践中考虑这个因素的影响。然而双耳受到 90dB(A)纯音与受到 90dB(A)全频带声音的影响是不一样的。纯音的能量集中在基底膜特定位置上，全频带声音的能量分布在整个基底膜上，纯音比全频带声音对听力的损伤更大。需要指出的是，噪声评价计量一般要进行“A”加权，虽然这种加权方法更适合于评价声压级较低的情况，并且在听觉最敏感的 4kHz 附近权重最大。

图 2.20　最长暴露时间随声压级变化的线性表示（上图）和对数表示（下图）（注意在欧洲自 2006 年 4 月起声压级上限有所调整）



2.5.2　听力保护

听力损失是隐伏和永久性的，当听力损伤可以测量到时往往为时已晚。因此，必须以积极主动的态度保护听力的灵敏度和分辨力。首先，要避免听过大的噪声。尽管声压级 90dB 视为听力损伤的阈值，但是，当噪声引起双耳振铃现象时，尤其是振铃持续时间较长时，即使声压级低于 90dB 也有可能损伤听力。

下面列举了一些潜在的可能损害听力的情况。

第一，用耳机听录音节目时，即使很小的耳机也有可能达到损伤听力的声压级。

第二，在演奏音乐时，不管使用声学乐器还是电子乐器，都有可能达到损伤听力的声压级，尤其是在声学环境活跃的小房间里进行演奏时。请参看第 6 章。

在这两种情况下，听音者都可以控制声压级，因此可以将声压级降低。然而，还有一种现象叫做听反射（见第 2.1.2 小节），当重放声音的响度大时听觉会降低其灵敏度。当这个现象与听阈暂时提高的现象结合时，会导致声压级盘旋式上升，即听音者“想要听得更清楚”，于是会增大响度，这会使声音变得更加暗淡，于是继续增大响度，如此循环下去。正确的解决方案是避免重放声的响度过大，但如果听觉系统的灵敏度已经降低了，应该暂时离开响度大的听音环境，使听觉的灵敏度得到恢复。

在有些情况下听音者不能对声源的声压级进行控制，例如乐队、迪斯科、夜总会以及电动工具产生的声音。在这种情况下，较好的方法是限制噪声暴露的剂量，或者更好的方法是实行听力保护措施。例如，可以与音乐会或迪斯科扬声器保持一个合理距离。听力的分辨力往往需要几天甚至几周的时间才能完全得到恢复，因此要避免每天都参加响度过大的音乐会或夜总会！每当要参加声压级大的活动时，作者通常使用小型耳塞保护听力，很多专业音响工程师也会有同样的做法。这种方法既不会引人注目还能适度地降低声压级 15dB ~ 20dB，使在噪声干扰的情况下还可以用正常讲话的音量进行谈话！这种耳塞还具有平直的频率衰减特性，因此不会过分严重地影响声音的平衡，它的成本还不到一张 CD 的价钱。对于响度很大的声音，例如电动工具产生的噪声，则需要采取更加严格的听力保护措施，例如使用降噪耳机。

我们的听力对于欣赏音乐、与人沟通交往都是至关重要而不可取代的。不管现在和将来，听力都应该受到保护。

2.6　听觉对声源的定位能力

我们利用双耳可以判断声音的方向，然而是如何判断的呢？由于双

耳位于头部的两侧，这样的构造对于不同方向的声源具有不同的声学效果。两只分离的耳朵会对声波产生两种现象：第一，声波到达双耳的时间不同而产生的双耳时间差（interaural time difference，缩写为 ITD）；第二，声波到达双耳的强度不同而产生的双耳声级差（interaural intensity difference，缩写为 IID）。这两种现象是截然不同的，下面分别进行介绍。

2.6.1 双耳时间差（ITD）

双耳时间差与头部形状有很大关系，图 2.21 所示为水平面上双耳与声源的相对位置关系。由于双耳的间距约为 18cm，因此声波到达双耳的距离一个较近，另一个较远，因此形成一定的时间差。当声源位于左侧时，左耳更早地接收到声音；当声源位于右侧时，右耳则更早地接收到声音。如果声源处于正前方或正后方，或中垂面上某一位置，声音将同时到达双耳。双耳时间差取决于声波分别到达双耳的距离差。在图 2.22 中，设双耳间距为 d，利用声源方位角可计算出声源到达双耳的时间差为：

$$\Delta t = \frac{d\sin\theta}{c}$$

其中，Δt 为双耳时间差（s），d 为双耳间距（m），θ 为声波入射方向与中垂面的夹角（弧度），c 为声速（m/s）。

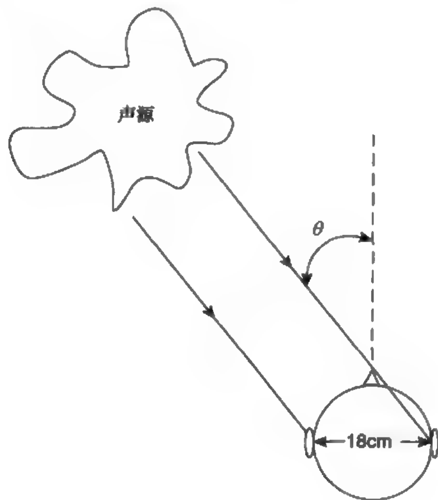
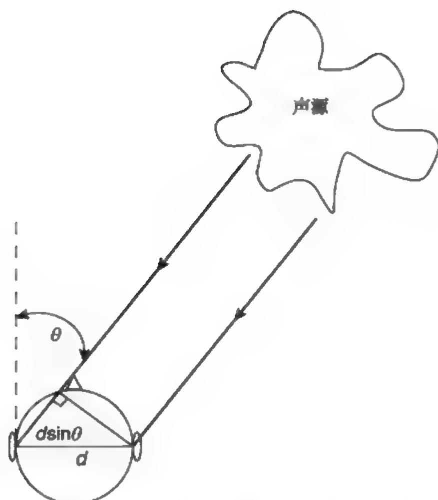


图 2.21 声源与头部的相对位置关系

图 2.22　双耳时间差的简单计算模型

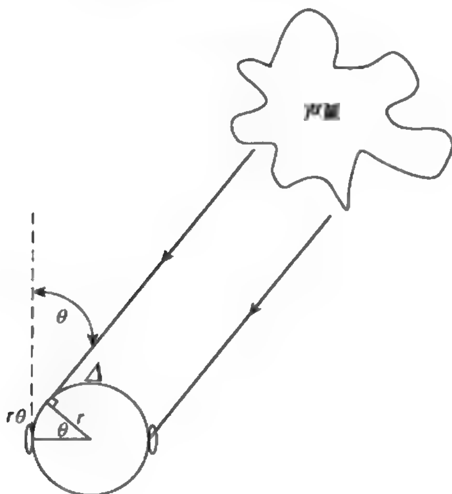


然而这个公式表达并不准确，因为它忽略了声波必须绕过头部才能到达双耳，因此它低估了双耳之间的延迟时间。声波在头部的绕射增大了延迟时间，如图 2.23 所示。假设头部为球状，则这个额外的延时是可以计算的。声波绕过头部的路程可以用如下关系计算：

$$\Delta d = r\theta$$

其中，Δd 为声波以一定角度入射绕过头部而产生的额外声程（m），r 为双耳间距的一半（m），θ 为声波入射角。

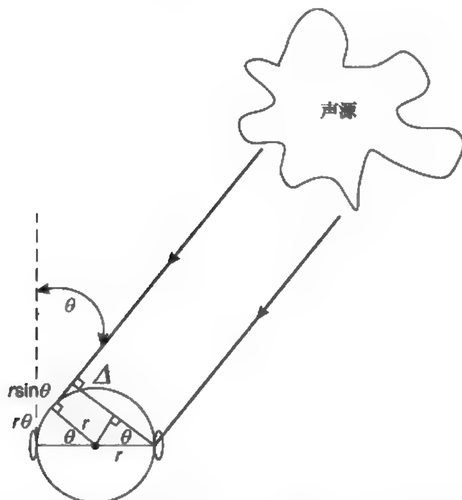
图 2.23　声波绕过头部对双耳时间差的影响



考虑到声波绕射产生的额外路程差，它是随半径 r 变化的函数，参看图 2.24，双耳时间差（ITD）更准确的表达式为：

$$ITD=\frac{r(\theta+\sin\theta)}{c} \tag{2.9}$$

图 2.24　双耳时间差的更准确计算模型



用上述公式可以计算最大双耳时间差，即声源入射角度为 90° 或弧度为 π/2 时的 ITD 为：

$$ITD_{\max}=\frac{0.09\times(\theta/2+\sin(\theta/2))}{344}=6.73\times10^{-4}\ \mathrm{s}(673\ \mu\mathrm{s})$$

虽然这个延迟时间非常短，但是时间差从该值到零的变化却能决定低频声波的方向。图 2.25 表示双耳时间差与入射角的关系，正延时与声源位于中垂面右侧相对应，负延时与声源位于中垂面左侧相对应。要注意的是，声源以相同的角度在前方和后方入射产生的时间差是一样的，也就是说，听觉还需要运用其他原理和方法才能判断声源是来自前方还是后方。当然，听觉能够利用双耳时间差判定声源方位的频率范围是有限的，这是因为听觉是通过双耳时间差来判断声波的相位差从而定位声源的方向。这就是说，双耳接收到的相位差是：

$$\varphi_{ITD}=2\pi f\cdot\frac{r(\theta+\sin\theta)}{c}$$

其中，φ_{ITD} 为双耳相位差，f 为频率（Hz）。

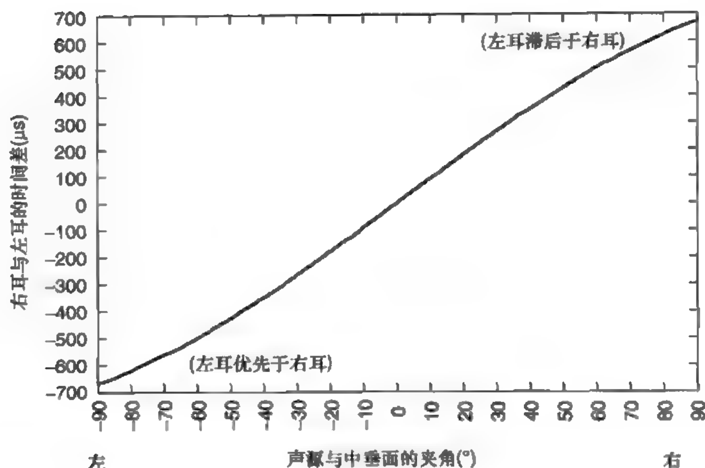


图 2.25　双耳时间差（ITD）与声波入射角的关系

当相位差大于弧度 π 或 180° 时，双耳无法分辨声源的方向，这是因为两种入射角度（一个在左侧另一个在右侧）都可能产生一个相同的相位差。因此，声波在一定角度入射时，听觉可以应用双耳时间差判定声源方位的上限频率为：

$$f_{max}(\theta)=\frac{c}{2\times0.09\times(\theta+\sin(\theta))}$$

当声波以 90° 入射时，上限频率为：

$$f_{max}(\theta=\pi/2)=\frac{c}{2\times0.09\times(\pi/2+\sin(\pi/2))}=743\text{Hz}$$

因此，当声源以 90° 入射时，可通过相位差分辨声源方向的最大频率为 743Hz。当声源入射角减小时，频率上限会随之增大。

2.6.2　双耳声级差（IID）

另外一种声源定位的方法是利用双耳声级差，它是由头部遮蔽效应产生的。头部对声波的遮蔽效应如图 2.26 所示，当声源位于中垂面时，到达双耳的声级是相等的；当声源偏离中垂面时，一只耳朵的声级逐渐减小，另一只耳朵的声级逐渐增大。头部遮蔽效应很难通过计算来衡量，有实验表明，两耳的强度比随着声源的入射角度以正弦方式在 0dB ~

20dB 变化，变化的多少与频率有关，如图 2.27 所示。

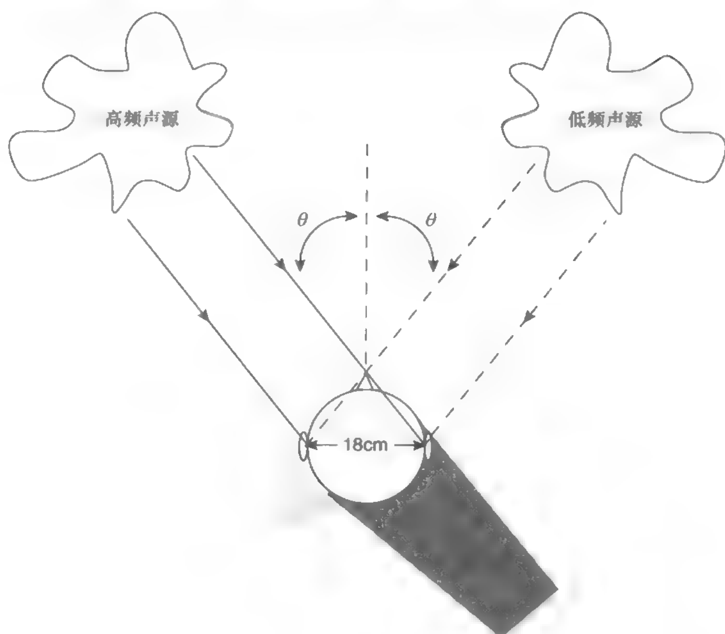


图 2.26 头部遮蔽效应引起的双耳声级差

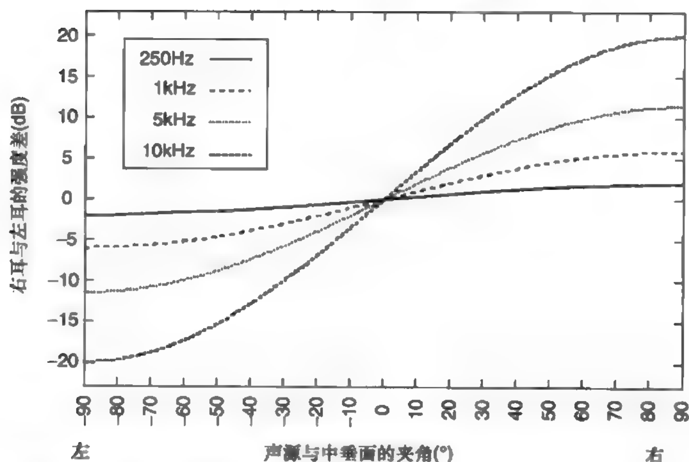


图 2.27 双耳声级差（IID）随声波入射角和频率变化的特性

然而正如第 1 章所述，当物体的尺寸达到声波波长的 2/3 （2/3λ）时，声波的散射和遮蔽现象才比较明显。由此可以认为，当物体的尺寸为声波波长的 1/3 （1/3λ）时才开始发生散射。这就是说，存在一个最小频率，当低于这个最小频率时，声级差对声源的定位作用不大。如果以头

部的尺寸大约为声波波长的 1/3 （1/3λ）确定这个最小频率，头部直径取 18cm，那么最小频率为：

$$f_{\min(\theta=\pi/2)}=\frac{1}{3}\left(\frac{c}{d}\right)=\frac{1}{3}\times\left(\frac{344}{0.18}\right)=637\,\text{Hz}$$

因此，双耳声级差主要对高频声音的定位起作用，而双耳时间差主要对低频声音的定位起作用。要注意的是，大约在 700Hz ~ 2.8kHz（700Hz 的 4 倍），听觉完成这两种定位方法的过渡。双耳对这两个频率之间的声音信号的定位能力不及对其他频率的定位。

2.6.3　耳廓和头部转动对定位的作用

上述听觉方向定位模型无法解释听觉如何判断声源来自前方还是后方，以及如何判断声源高度的问题。事实上，听觉有两种方法可以解决这些问题。

首先，利用外耳对声波的作用来定位声源的角度和方向。这是由于辐射到耳廓的声波经过反射传输到外耳道，耳廓的反射声存在一定的延时，延时量虽然很小，但是却非常有效，它与直接进入外耳道的声波产生干涉形成梳状滤波效应。由于声音到达的方向不同，反射声与直达声的时间差会发生变化，因此形成了一种与声源方向相关的频谱特性，听觉系统据此判断声音的空间方位。反射声与直达声的时间差非常小，因此这种定位方法主要适用于高频信号，即高于 5kHz 的频率范围。这种定位效果因人而异。因为人与人之间耳廓的形状是不同的，并且随着年龄的增大，存在一个学习过程，使这种利用耳廓定位声源的能力逐渐增强。因此，在彻底改变头部声学形状的时候，例如把长头发剪短，听觉定位会变得混乱。此外，我们还发现，当我们听用“别人的耳朵”录制的节目时，我们定位声音的能力会变得不同，因为声波在耳廓的干涉模式与自己的有所不同。有时候听到的声音不如自己耳朵录制的，有时候比自己耳朵录制的声音还要好。

第二，头部转动是解决声像定位不明确的一种有效方法。听一个声音时，我们想要判断它的方向，于是将头转向声源，甚至想让声源位于我们的正前方，此时所有的时间差和强度差将为零。头部转动将改变声

源的入射方向，声源方向的改变量取决于声源与听音者之间的相对位置关系。因此，对于来自后方、前方或上方的声源，头部摆动时声源的移动方向是不一样的。这也是使用耳机监听会产生头中定位效应的原因。因为声源的方位不随头部的摆动而变化，因此听觉本能地判定声源位于头内，而不是在头外。产生头中定位的另一个原因是耳机重放不能很好地模拟自然听音时的双耳信号。实验结果表明，当耳机重放能够模拟头部效应，并且当头部转动时能保持声源的方向不跟随头部时，重放声的声像定位效果会更加令人信服。

2.6.4 双耳时间差和声级差的共同作用

由于强度差和时间差都用于感知声源的方位，因此人们认为这些听觉的感知机构可能位于大脑内相临近的部位，而且互相联系在一起。如果是这样的话，人们自然会认为大脑对双耳时间差和强度差的诠释在某些方面是相同的，使得大脑可能会错误地把强度差理解成时间差，反之亦然。这就使得一种定位信息可能抵消另外一种定位信息，例如强度差可以抵消时间差。这种现象实际上是存在的，称做双耳时间差和强度差的补偿关系。从定位效果上看，在一定的限度内，双耳时间差可以用一个适量的双耳强度差进行补偿，如图 2.28 所示，它们有一些有趣的特征。首先，正如所预料的，只有一定范围内的时间差才能用强度差补偿，

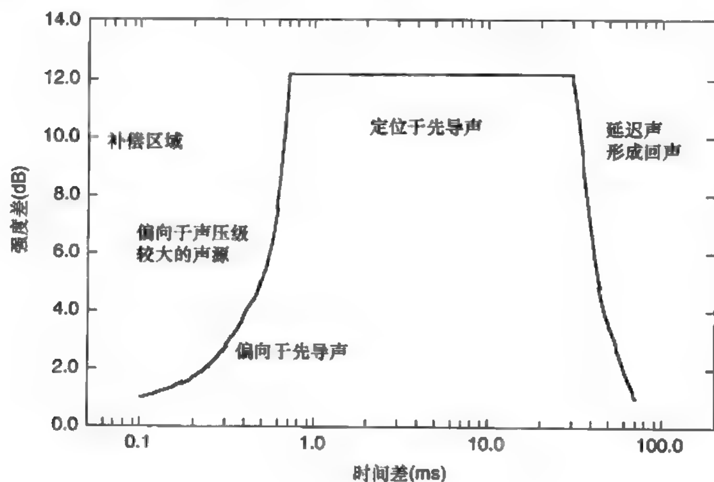


图 2.28 时间差和强度差的补偿关系（Madsen 1990）

最大可补偿的时间差约为 673μs。当超出这个时间差时，较小强度差的变化并不能改变声像定位，听觉将一直定位于先导声的方向。听觉定位于先导声的现象一般发生在 673μs ~ 30ms 的范围内，但是，当延迟声比先导声声压级大 12dB 以上时，听觉定位将受到延迟声的影响。当时间差大于 30ms 时，延迟声将可能形成回声，听音者此时可以分辨出延迟声和先导声。以上现象具有两层含义：第一，听觉可以单纯通过时间差或强度差来获得方位信息；第二，一旦声源延时超过 700μs 时，如果不考虑先导声和延迟声相对声压级的影响，听觉将定位于先导声。如果先导声的声压级明显低于延迟声，则听觉定位于先导声的现象将消失。

2.6.5　哈斯效应

双耳时间差和强度差补偿效应的第二种情况（见图 2.28）也称为哈斯效应或优先效应，是以对这个效应进行实验后得到定量结果的实验者名字命名的。哈斯效应主要包含以下两个含义：

（1）延时在 30ms 以内时，听觉将定位于先导声，而不能感觉到延迟声的存在；

（2）听觉一般不能察觉 30ms 以内的反射声，如果反射声在 30ms 以后到达，反射声可能以回声的形式被听到。

这个研究结果对演播室、音乐厅和扩声系统方面的工作有非常重要的指导意义。基本要求是保证早期反射声在 30ms 以内到达观众区，以避免这些反射声形成回声。实际上亲切感较强的音乐厅，其反射声的延时大多都在 20ms 以内。在扩声系统中，扬声器的输出经常相对于声源有一定的延时，由于哈斯效应的存在，在扬声器响度不太大的情况下，听音者将定位于实际声源方向。

2.6.6　立体声重放

基于听觉对声源的定位机理，在立体声重放中，通常只利用一对扬声器或耳机就可以获得声像定位感。立体声制式基本上有 3 种，前两种是基于只提供听觉系统两种主要定位信息之一的理念，也就是利用强度

差或时间差定位。还可以借助双耳时间差和强度差的补偿性形成既有时间差信息又有强度差信息的立体声制式。这两种立体声制式分别说明如下：

（1）时间差式立体声：图 2.29 所示的拾音系统包含两只具有一定间隔的全指向性传声器，与声源保持一定距离。由于传声器的间隔小，因此拾取的声压级是一样的，只是存在一定的时间差。所以这两支传声器拾取的信号在进行重放时，只包含时间差定位信息。

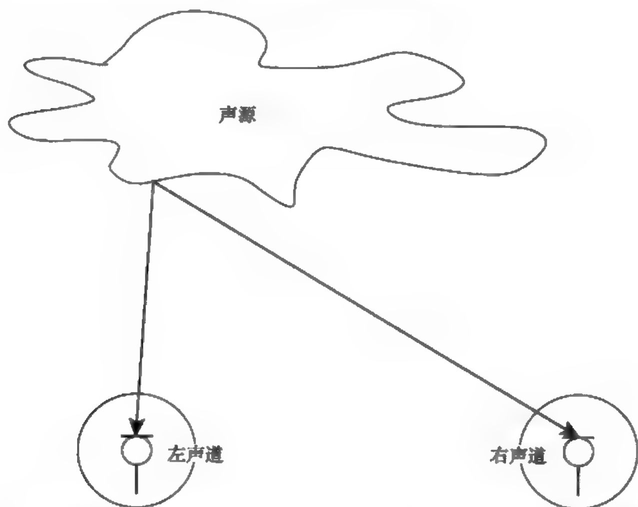


图 2.29 时间差式立体声拾音

（2）声级差式立体声：图 2.30 所示的拾音系统包含两只主轴形成一

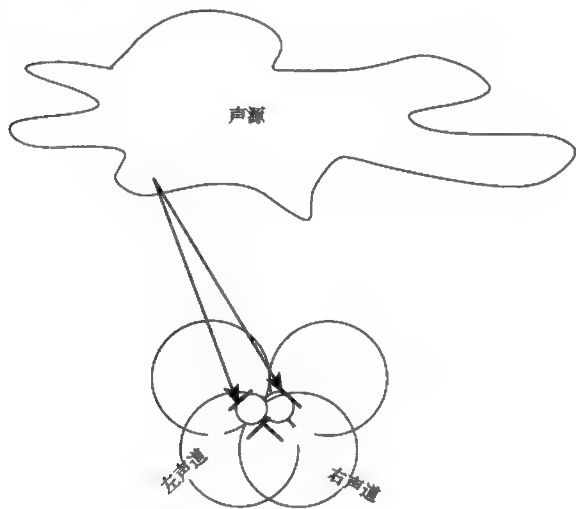


图 2.30 声级差式立体声拾音

定夹角、放置在同一点的指向性传声器，分别指向声源的左侧和右侧。由于传声器放置在同一点，因此拾取的声音不存在时间差。由于传声器具有一定的指向性，因此两支传声器拾取的信号强度是不一样的。所以这两支传声器拾取的信号在进行重放时，只包含声级差定位信息。声级差式立体声拾音技术在流行音乐制作中应用非常广泛，包括调整调音台上声像电位器以控制两个声道输出不同强度的信号，从而确定某一声源在立体声声像舞台上的位置，如图 2.31 所示。

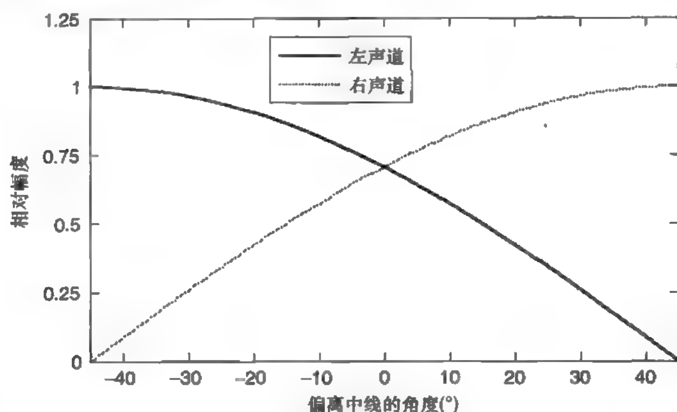


图 2.31　调音台上声像电位器对两个声道信号强度的分配

上述两种方法的区别在于记录原始声源的方法不同，而与声音的重放方式无关，所以究竟应该采用哪种技术进行录音取决于制作人和录音师的需要。通过采用不同的传声器拾音技术，还可以将这两种定位信息结合起来使用。例如，将指向性传声器以一定间距放置进行拾音，这样拾取的立体声将同时包含时间差和声级差信息。但是这种拾音技术容易产生不真实的定位信息，使听觉定位不确定，因此对于这种拾音方式，需要在不真实的和所需的信息之间取得一定平衡以获得良好的声像定位，而做到这一点是比较困难的。

（3）仿真头式立体声（binaural stereo）：第 3 种提供定位信息的方法是利用仿真头立体声技术。如图 2.32 所示，这个系统需要将两个性能完全相同的全指向性传声器放置在真人或更多的是仿真头的外耳道内，然后将拾取的信号用耳机重放。两只传声器的间距与双耳间距相同，放置在能够模拟人头甚至肩部对声波遮蔽效应的物体上。这意味着随着声源方位的改变，传声器不仅能拾取到时间差信息，而且还能拾取到强度差

信息，声音的重放效果非常逼真自然。但是，这种立体声要用耳机重放，如果用扬声器重放则会发生声道串扰，容易产生不真实的定位信息，破坏重放效果。请注意，如果用耳机重放扬声器立体声节目时，也会发生类似的现象，因为此时声道串扰是必要的声像定位信息，而耳机重放就不存在这种串扰，这也是用耳机重放传统立体声产生头中定位效应的原因。

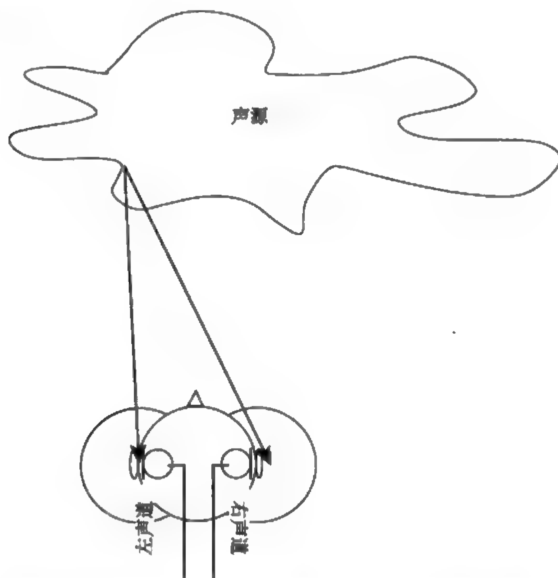


图 2.32 仿真头立体声录音

在立体声重放中要考虑和解决的主要问题是由听音环境产生的错误定位信息，因为扬声器及其所在的听音房间会提供关于扬声器位置的听觉信息，而这与原来的录音环境没有任何关系。更多有关听觉定位和立体声重放的问题可以查阅 Black(1997) 和 Rumsey(2001) 的资料。

参考文献

von Békésy (1960). Experiments in Hearing. New York: McGraw-Hill.

Blauert, J. (1997). Spatial Hearing: the Psychophysics of Human Sound Location, revised edn. Cambridge, MIT press.

Fletcher, H. and Munson, W. (1933). Loudness, its measurement and

calculation. Journal of the Acoustical Society of America, 5, 82–108.

Glasberg, B.R. and Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data, Hearing Research, 47, 103–138.

Gulick, L.W. (1971). Hearing: Physiology and Psychophysics. Oxford: Oxford University Press, pp. 188–189.

Madsen, E.R. (1990). In The Science of Sound (T.D. Rossing, ed.), 2nd edition, Addison Wesley, p. 500.

Pickles, J.O. (1982). An Introduction to the Physiology of Hearing. London: Academic Press.

Rumsey, F. (2001). Spatial Audio. Oxford: Focal Press.

Scharf, B. (1970). Critical bands. In Foundations of Modern Auditory Theory, Vol. 1 (J.V. Tobias, ed.), pp. 159–202. London: Academic Press.

第 3 章　乐音与和声的物理和生理基础

3.1　乐音

所有不同文化背景的音乐都是以那些可以发出不同音高乐音（译者注：这里乐音是指单音或音符）的乐器（包括人声）的使用为基础的。来自不同文化的音乐所利用的音高组合或许是独特的，但是人类对音高感知的心理声学基础却是相同的。这一章我们将探讨听觉能够感知其音高的音符的声学特性以及音高感知的心理声学，然后探讨在西方音乐中使用的各种调律系统的声学和心理声学。

由于存在很多不同的记谱法，音乐的音高常被混淆。在本书中我们采用的记谱方式是用 A4 表示钢琴键盘中央 C 之上的音符 A，在音符 B 和 C 之间的数字编号会有改变，总是用大写字母表示音符。因此中央 C 就是 C4，而中央 C 下面的音符 B 就是 B3，以此类推。88 键钢琴键盘的最低音记为 A0，因为它位于中央 C 下面第 4 个 A 的位置，同理键盘最高音是 C8。这种记谱方式与键盘的对应关系如图 3.21 所示。

3.1.1　乐音及其基频

当我们聆听某个乐器演奏一个音符时，我们会清晰准确地感知到音高，这是因为乐器振动产生了具有一定周期的声波。例如，图 3.1 所示的这些声压波形，它们是用一只传声器分别拾取的小提琴、小号、长笛和双簧管 4 种管弦乐器发出的 A4 音。我们注意到在每种情况下，波形是有规律且重复的，或者说波形具有周期性（见第 1 章）。每一个重复部分可以看作是一个周期，每一个周期所持续的时长称为“基本周期”或者波形的“周期”。一秒钟内周期的个数就是基频，用赫兹或 Hz 表示。基频通常记为 f_0，这将贯穿在本书其余章节中使用。因此任何波形的 f_0 都可以从波形周期计算出来，即：

$$f_0 = \frac{1}{T} \tag{3.1}$$

其中，T 为波形周期（s）。同时，周期也可以通过频率 f_0 得出：

$$T = \frac{1}{f_0} \tag{3.2}$$

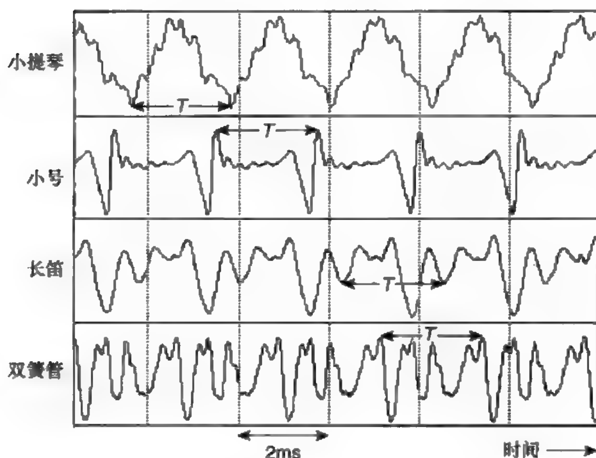


图 3.1　小提琴、小号、长笛和双簧管演奏 A4（440Hz）的声压波形（T 为周期）

例 3.1　试计算音符 G5 的周期，并找到用钢琴演奏波形周期为 5.41ms 的琴键。

由图 3.21 可知，G5 的基频为 784.0Hz，因此由式（3.2）可得出它的周期为：

$$T_{G5}=\frac{1}{784.0}=1.276\times10^{-3}\text{s}=1.276\text{ms}$$

波形周期为 5.41ms 的基频也可以由式（3.1）得出：

$$f_0=\frac{1}{5.41\times10^{-3}}=184.8\text{ Hz}$$

琴键上基频最接近 184.8Hz 的音符是 F#3（见图 3.21）。

从图 3.1 可以看出，小提琴的基频可以通过波形图上任意一个周期计算出来。周期是从某个周期的任意一点到下一个周期的同一点之间的时长，例如可以选择波峰、波谷或零振幅点。图中小提琴波形的周期 T_{violin} 是通过测量波谷之间时长得到的，实验室里测得的数据是 2.27ms，用式（3.1）可计算出基频为：

$$f_0=\frac{1}{(2.27\times10^{-3})}=440.5\text{Hz}$$

这个频率接近 A4 调音的参考频率 440Hz，如图 3.21 所示。如果一个音符不是在空弦上弹奏的，由于调音误差、音调抑扬变化和颤音技巧的应用，会使测量的基频和图 3.21 所示的参考基频之间存在一定差异。在实际测量中一般采用在不同时段多次测量取平均的方法测量基频 f_0。

3.1.2 乐音及其谐波

图 3.1 也显示了其他乐器演奏音符 A4 所产生的声压波形。尽管它们之间的周期和基频数值是相近的，但波形却截然不同。每一个乐器发出的音高都是 A4，但每种乐器的音色是不同的，这与其声压波形的细节有关。听众就是通过这些波形的细节去识别不同的乐器如小提琴、单簧管或双簧管。这是因为乐器振动产生的声压作用于听音者的鼓膜，从而决定了听觉基底膜上振动形态的建立，听觉就是根据这个振动形态对声源的频率成分进行分析（见第 2 章）。不管听到的音高是否相同，当比较不同声源激发的基底膜振动形态发生变化时，例如小提琴和单簧管，听觉就能感觉到音色的不同，

每种乐器在它的频谱上都有一系列频率分量（见第 1 章），听觉可以根据其频谱辨识不同的乐器。这些频率分量的结构可以看作是乐器所具

有的特定音色的“秘方”。图 3.1 所示为 4 种不同管弦乐器演奏同一个音符的声压波形，图 3.2 为相应的振幅频谱。从图可以看到，每个音符的波形是不同的，这样就形成了不同的频谱结构。虽然每个音的音高都是 A4，但不同乐器的音色是不同的。有音高感的乐器如小提琴、双簧管、单簧管和小号的乐音是由谐波成分组成的，其频率是基频的整数倍（1 倍、2 倍、3 倍、4 倍、5 倍等），详见第 1 章。因此，图 3.1 中基频为 440.5Hz 的小提琴的声压波形的频率成分为 440.5Hz(1×440.5Hz)、881.0Hz(2×440.5Hz)、1 321.5Hz(3×440.5Hz)、1 762Hz(4×440.5Hz)、2 202.5Hz(5×440.5Hz) 等。由图 3.2 可以看出，这些频率和频谱中的峰值频率相对应（见第 1 章）。这些谐波通常用它们的谐波次数命名，也就是用其频率对基频的倍数来命名。

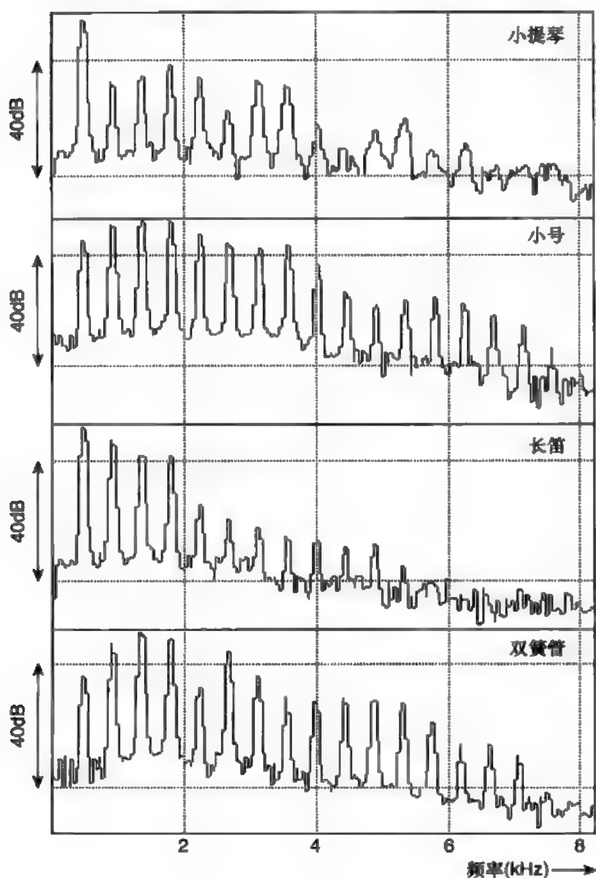


图 3.2 图 3.1 中小提琴、小号、长笛和双簧管分别演奏 A4 的频谱图

另一个早期用于描述周期波频率分量的名词是泛音，现在仍然为许多作者所采用。第一个高于基频的分量称为第1次泛音，也就是2次谐波；第2次泛音就是3次谐波，以此类推。表3.1总结了基频、泛音和谐波三者之间从1～10的倍数关系。

表3.1 周期性波形的前10个频率分量的泛音、谐波与基频之间的关系

整数倍（N）	泛音系列（第N-1次）	谐波系列（第N次）	分量频率（Hz）
1	基频（f_0）	第1次谐波	$1f_0$
2	第1次泛音	第2次谐波	$2f_0$
3	第2次泛音	第3次谐波	$3f_0$
4	第3次泛音	第4次谐波	$4f_0$
5	第4次泛音	第5次谐波	$5f_0$
6	第5次泛音	第6次谐波	$6f_0$
7	第6次泛音	第7次谐波	$7f_0$
8	第7次泛音	第8次谐波	$8f_0$
9	第8次泛音	第9次谐波	$9f_0$
10	第9次泛音	第10次谐波	$10f_0$

例3.2 试计算音符基频为101Hz的第4次谐波频率，并求出音符基频为120Hz的第6次泛音频率。

第4次谐波频率为$4f_0$：（4×101）=404Hz

第6次泛音频率为$7f_0$：（7×120）=840Hz

尽管很多乐器的谐波结构都受到了乐器本身构造的限制，但是理论上任何乐器输出的谐波次数都是没有上限的。上限频率就是听觉系统能够听到的最大频率，大约为16 000Hz（见第2章）。因此乐器在弹奏中央C之上的音符A时，它的基频是440Hz，理论上在人耳听觉范围内，它的谐波次数最大为36（=16 000/440）次。如果乐器弹奏的是高八度音，那么基频是880Hz，理论上它的谐波次数最大为18（=16 000/880）次。尽管上限频率通常会受到乐器本身声学条件的限制，使这个上限一般是在人耳听力上限频率之内，但是对于电子合成器来说，可以很容易地将高次谐波扩展到听觉上限频率以上。这是一个有趣而重要的现象。

3.1.3 谐波之间的音程关系

从声学角度来说，一个能被感知音高的声音都是由称为谐波的基频

整数倍的频率成分组成，每个谐波都是一个正弦波。由于听觉系统是根据频率成分分析声音的，因此这些谐波对理解周期性声波的分析与合成是非常重要的，而且研究各个谐波之间的音乐关系，是西方音乐和声发展的核心内容。这一系列谐波的频率比见表 3.1 所示，第 1 次到第 10 次谐波的音程关系、频率比、相对键盘中央 C 的位置以及五线谱位置如图 3.3 所示。现代键盘上的音程关系（除倍频程以外）只是近似于上述的音程关系，这与所使用的调律系统有关，这部分内容将在第 3.4 节介绍。

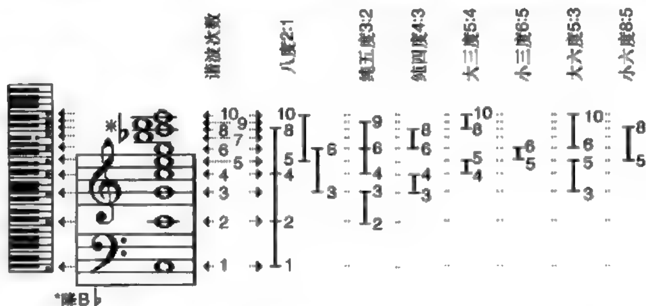


图 3.3　音符 C3 的前 10 次谐波的频率比和音程关系以及在键盘和五线谱的位置

在一系列自然谐波中，相邻谐波的音程开始于基频或者第一次谐波，在图 3.3 中用五线谱和键盘上琴键表示出来，它们依次是八度（2 ∶ 1）、纯五度（3 ∶ 2）、纯四度（4 ∶ 3）、大三度（5 ∶ 4）、小三度（6 ∶ 5）、减三度（7 ∶ 6）、增二度（8 ∶ 7）、大全音（9 ∶ 8）和小全音（10 ∶ 9）（译者注：图中只表示出其中部分音程）。非邻近谐波之间的音程频率比也可以从图中得出。例如，第 4 次谐波与基频之间的音程是两个倍频程，而频率比是 4 ∶ 1，相当于两次加倍的关系。同理，3 个倍频程的频率比是 8 ∶ 1，1 个倍频程加上 1 个五度的频率比是 3 ∶ 1。

其他常用的音程可以从这些音程中找出（图 3.15 所示为一个八度内的所有音程）。我们先用已知的结果说明音程频率比的计算方法。从纯五度的频率比（3 ∶ 2）可以推得纯四度的频率比（4 ∶ 3），因为它们的音程加在一起构成了一个倍频程（2 ∶ 1），即从 C 到 G（纯五度）和从 G 到 C（纯四度）。请记住音程从频率关系上看就是频率比，因此音程的加减等效于频率比的乘除。一个纯四度的频率比可以通过一个八度的频率比除以一个纯五度的频率比得到，或者将纯五度的冠音下行一个八度或根音上行一个八度得到，即：

$$\text{纯四度的频率比}=\frac{2}{1}\div\frac{3}{2}=\frac{2}{1}\times\frac{2}{3}=\frac{4}{3}$$

另外两个常见的音程是大六度和小六度。它们的频率比可以分别从小三度和大三度计算得出，因为它们都可以结合成一个完整的八度。

例 3.3 试找出大、小六度的频率比。已知八度的频率比为（2 ∶ 1），小三度的频率比为（6 ∶ 5），大三度的频率比为（5 ∶ 4）。

因为一个大六度和一个小三度构成一个八度，所以：

$$\text{大六度的频率比}=\frac{2}{1}\div\frac{6}{5}=\frac{2}{1}\times\frac{5}{6}=\frac{10}{6}=\frac{5}{3}$$

又因为一个小六度和一个大三度构成一个八度，所以：

$$\text{小六度的频率比}=\frac{2}{1}\div\frac{5}{4}=\frac{2}{1}\times\frac{4}{5}=\frac{8}{5}$$

这些频率比也可以从音程和谐波系列中推出。由图 3.3 可知，大六度音程是第 5 次谐波和第 3 次谐波之间的音程，在这个例子中是 G4 和 E5，所以频率比是 5 ∶ 3。同样地，小六度音程是第 5 次谐波和第 8 次谐波之间的音程，两个音是 E5 和 C6，所以小六度音程的频率比是 8 ∶ 5。这些关于音符和谐波系列方面的知识从音乐方面和声学方面都是非常有用的，是所有铜管乐手和了解变化音栓（见 5.4 节）的管风琴演奏者非常了解的。

图 3.4 表示 A3（f_0=220Hz）的前 10 个谐波频率分别在线性坐标轴和对数坐标轴上的位置。请注意，谐波之间的间隔在线性坐标轴上是相等的，而

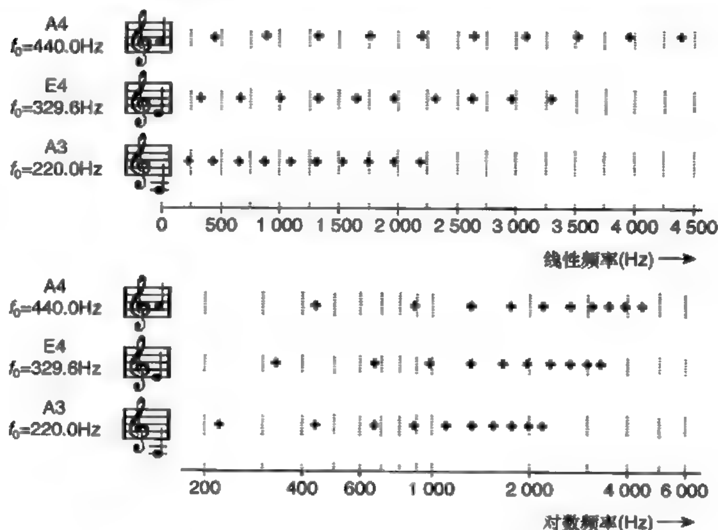


图 3.4 在线性坐标（上图）和对数坐标（下图）上的 A3（f_0=220Hz）、E4（f_0=329.6Hz）和 A4（f_0=440.0Hz）的前 10 次谐波频率位置

在对数坐标轴上间距会随着频率的增大变得越来越小。尽管看上去对数坐标轴上谐波的分布比线性坐标轴复杂，但是要注意大自然总是选择最高效的处理方式。对于基频不同的音符，例如 E4(f_0=329.6Hz)和 A4(f_0=440.0Hz)，其谐波分布形式在对数坐标轴上是一致的，但在线性坐标轴上却不相同。上述特性和音色感知有很大关系，这将在第 5 章做进一步讨论。

由于听觉系统的基底膜部位和对数频率的分布相一致，因而对数频率坐标最能代表听觉系统对固定音高乐器音符的各谐波成分的感觉加权特性。

在听觉感知的研究中使用对数频率坐标的好处在于，在分析频率比相同的频率分量时，听觉系统可以给出相同的感觉权重。图 3.5 所示为一些常用的音程标注在对数频率轴上，一直延伸到听觉的上限频率。为方便起见，它们都以 A1(f_0=55Hz) 为起点，而且可以用任何其他音程建立类似的图表。值得注意的是在对数坐标上相同音程之间的间距是相等的。用尺子可以很容易地验证这一点。例如，在 X 轴上从 100Hz 开始测量倍频程之间的距离(例如，100Hz 与 200Hz、200Hz 和 400Hz、400Hz 和 800Hz 之间等)，并与图 3.5 中倍频程关系图上两点之间的距离进行比较，看它们是否相等。在对数频率轴上，无论是 2 ∶ 1、3 ∶ 2 或 4 ∶ 3 的音程关系，轴上任意一段距离都等于一个常数，无论位于轴线上的什么位置。可以制作一个用音程校准过的音程标尺，用于在对数频率轴上方便地找到两个一定间隔音符的频率比。然而，这样的校准必须考虑常用的音程，例如八度 (2 ∶ 1)、纯五度 (3 ∶ 2)、大六度(5 ∶ 3)等。如果纯五度的音程距离叠加在纯四度的音程距离上，那么将得到一个八度的音程距离，因为八度等于一个纯五度加上一个纯四度。同样地，如果一个大六度叠加了一个小三度，那么也将得到一个八度音程，因为一个大六度加上一个小三度等于一个八度 (见例 3.3)。

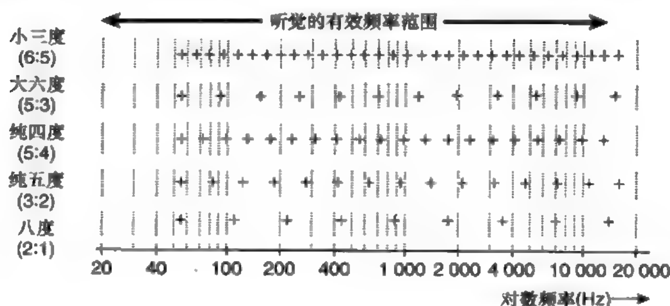


图3.5　由 A1(f_0=55Hz) 向上构成的八度、纯五度、纯四度、大六度和小三度在对数频率轴上的分布

如果在对数频率轴上把任意一个音符加倍（或者减半），那么在听觉上就相当于把这个音程提升（或降低）一个八度，乘以 3/2（或除以 3/2）在听觉上相当于把这个音程提升（或降低）了一个纯五度，以此类推。我们感知到的所有音程关系（八度、纯五度、纯四度、大三度等）不管出现在哪个频率范围都会很相似。例如，两个音符构成的大六度不管是用两个低音提琴演奏还是用两只长笛吹奏，它们在音程上所带来的感觉是相似的。以这种方式，听觉频率分析机理的对数本质为我们理解音程和音高的主观感觉奠定了基础。

对于没有固定音高（但又有音高感，见下文说明）的声音，如在英文单词 sea 里的“ss”（见图 3.9），它的声压波形并不呈现有规律的重复，而是随时间随机变化，因此不是周期信号。这样的波形称为“非周期性”波形（见第 1 章）。这类声波包含的频率成分不再是某个频率的整数倍，因此不包含谐波成分，而是包含了所有频率，这样的频谱结构称为“连续”频谱。图 3.6 所示为小军鼓鼓刷音的非周期性声压波形及其频谱。

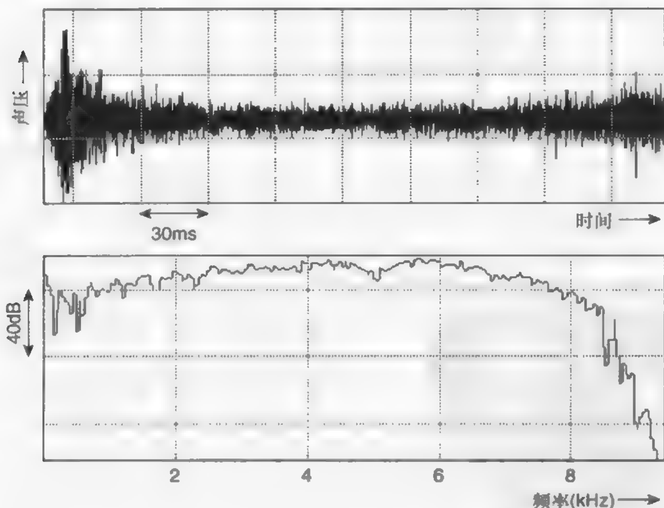


图 3.6　小军鼓鼓刷音的声压波形（上图）和频谱（下图）

3.2　音高感知理论

音高的感知是听觉感受具有调性的音乐的基础。熟悉目前的音高感

知理论以及心理声学的其他方面知识能够帮助我们较好地理解音乐方面的许多事物，如调音、音调、绝对音高、颤音、全新声音的电子合成以及音高幻觉现象（见第 5 章）。

音高是听觉对声音高低的感觉。美国国家标准协会（1960）给出的音高正式定义如下："音高是听觉对声音在音阶上按从低到高顺序排列的感知属性"。因为音高的测量需要听音者做出主观感觉的判断，因此具有"主观性"。这和在实验室里的物理测量方法是完全不同的。例如，测量一个音符的基频，它就属于"客观测量"。

总之，时间上周期性变化的声压波形被听觉系统接收后会产生一定的音高感，而听觉不对时间上非周期性变化的声压波形产生音高感知。波形和具有音高的频谱以及不具有音高的频谱之间的关系总结在表 3.2 中，这两种情况的例子见图 3.2 和图 3.6。在考虑信号的时间（波形）和频率（频谱）方面的特性时，广泛采用"时域"和"频域"这两个名词。

表 3.2　　有音高和无音高声音的波形和频谱特点

	有音高	无音高
波形 （时域）	周期性的 （有规律的重复）	非周期性的 （没有规律不重复）
频谱 （频域）	线状谱 （谐波成分）	连续谱 （无谐波成分）

声音的音高是随着其基频 f_0 的变化而改变的。基频 f_0 越大音高越高，反之亦然。尽管音高和基频 f_0 可以分别通过主观和客观测量得到，并分别以音符的高低为尺度以及 Hz 为测量单位，音高的主观测量结果也可以用 Hz 表示。这可以通过不断切换一个声音和频率可变的正弦波，让听音者对两个声音的音高进行比较而得到。听音者可以调整正弦波的频率直到他感知两个声音的音高相等，该声音的音高就是正弦波的频率大小。

"部位学说（place theory）"和"时间论（temporal theory）"是解释听觉感知音高的两个基本理论，它们可以解释人类的听觉系统如何定位和跟踪所接收的声音基频的变化。下文将对"部位学说"和"时间论"分别进行介绍，并说明它们在解释音高感知现象中还存在的局

限性。

3.2.1　音高感知的部位学说

音高感知的部位学说与基底膜的频率分析特性有直接的关系，输入的声音信号不同频率分量刺激基底膜的不同部位。毛细胞在基底膜的每个位置都会进行神经放电，并且刺激与输入声源的频率成分相对应的神经细胞和大脑高级中枢。假设声音中存在所有的谐波成分，下面是基于输入信号各频率成分的刺激部位分析找到基频 f_0，并使听觉神经和大脑的高级中枢能够对频率成分进行高级处理的几种可能性。

（1）方法 1：直接找出基频 f_0 本身的大小。

（2）方法 2：找出相邻谐波之间的最小频率差。如果所有的谐波都存在，那么根据定义第（n+1）和第 n 次谐波之间的频率差为：

$$(n+1)f_0-nf_0=nf_0+f_0-nf_0=f_0$$

其中，n 为 1，2，3，4，…。

（3）方法 3：找出所有频率成分的最大公因数（能够整除所有频率的最大整数）。表 3.3 所示为一个基频为 100Hz 的声音的前 10 次谐波频率以及分别除以 1 ~ 10 的整数后得到的数值，寻找在所有频率中都存在的最大数值。这些谐波频率排列在最左边的一列中（听觉部位分析的结果），其他列为每个谐频与整数（m=2 ~ 10）相除后的结果。最大公因数就是可能出现在表中每一行的最大数值，包括谐波本身（f_0÷1）即（m=1）在内。在这个例子中是 100Hz，听觉的音高感知就由这个频率决定。此外，有趣的是如果表格向右一直扩展下去，那么每一行中都会出现基频（100Hz）对应的那一行的数值。这是因为根据定义，每个谐波频率都为 100Hz 的整数倍（n），所有在 100Hz 这一行中出现的数值都是 100Hz 与整数（m）相除的结果，因此 100Hz 这一行中所有的数值都可以通过将谐波频率除以整数（m×n）得到。这些频率（50Hz、33Hz、25Hz、20Hz 等）的谐波也包含了所有给出的频率成分，因此被称为“次谐波（sub-harmonics）”。这就是为什么要用最大公因数的原因。

表 3.3　　找出 f_0=100Hz 的声音的前 10 个谐波的最大公因数的处理方法

部位分析	高级处理								
$n \times f_0$(÷1)(Hz)	÷2(Hz)	÷3(Hz)	÷4(Hz)	÷5(Hz)	÷6(Hz)	÷7(Hz)	÷8(Hz)	÷9(Hz)	÷10(Hz)
100	50.00	33.33	25.00	20.00	16.67	14.29	12.50	11.11	10.00
200	100.0	66.67	50.00	40.00	33.33	28.57	25.00	22.22	20.00
300	150.0	100.0	75.00	60.00	50.00	42.86	37.30	33.33	30.00
400	200.0	133.3	100.0	80.00	66.67	57.14	50.00	44.44	40.00
500	250.0	166.7	125.0	100.0	83.33	71.43	62.50	55.56	50.00
600	300.0	200.0	150.0	120.0	100.0	85.71	75.00	66.67	60.00
700	350.0	233.3	175.0	140.0	116.7	100.0	87.50	77.78	70.00
800	400.0	266.7	200.0	160.0	133.3	114.3	100.0	88.89	80.00
900	450.0	300.0	225.0	180.0	150.0	128.6	112.5	100.0	90.00
1 000	500.0	333.3	250.0	200.0	166.7	142.9	125.0	111.1	100.0

最早版本的部位学说指出，声音的音高对应于其最低频率成分（也就是基频）所刺激的基底膜部位（上述“方法 1”）。这个假设成立的前提条件是声音中总存在基频，并且该理论被欧姆在其第二定律即欧姆声学定律中概括为（他的第一定律是电气工程的基础，即：电压 = 电流 × 电阻）：“只有当声波在某一频率上具有一定能量时，那个频率的音高才能被听到”。

下面这个理论是通过严谨的实验得出的，当时人们已经能够合成给定频谱的声音信号。Schouten（1940 年）证明了去除基频成分后，脉冲波的音高依然保持不变，由此可以证明以下两点：（1）在感知音高时基频不是必须存在的；（2）最低的频率成分不是感知音高的基本依据。当基频缺失时，二次谐波是最低频率成分，音高不会因此而提升一个倍频程。这个实验被称为“缺失基频现象”，它说明了上述“方法 1”不能全面解释人类对音高的感知现象。

由于部位学说提出谐波也存在对应的刺激部位，不管基频是否存在，因此，只要声音中存在一些相邻的谐波成分，“方法 2”似乎对音高感知更加具有可行性。对大部分的乐音来说，确实存在相邻的谐波。然而，研究人员总是致力于寻求验证心理声学理论的方法。在验证音高感知理论方面，他们努力寻找一些声音，这些声音的音高感知现象不能用现有的理论解释。这些声音通常是通过电子手段产生的，可以对其频率成分和时间特性进行控制。图 3.7 所示为一个声音信号的理想化频谱，它只包含奇次谐波（$1f_0$、$3f_0$、$5f_0$……），相邻谐波的频率间隔分别为 f_0、$2f_0$、

$2f_0$、$2f_0$ 等，谐波之间的最小间距就是基频 f_0，它是感知音高的基本信息。但是，如果基频成分被去除（假定去除图 3.7 中所示的 f_0 点），音高是不会改变的。然而去除 f_0 点后，相邻谐波之间的间距为 $3f_0$、$2f_0$、$2f_0$、$2f_0$ 等，最小间距为 $2f_0$，但是音高并没有提升一个倍频程。

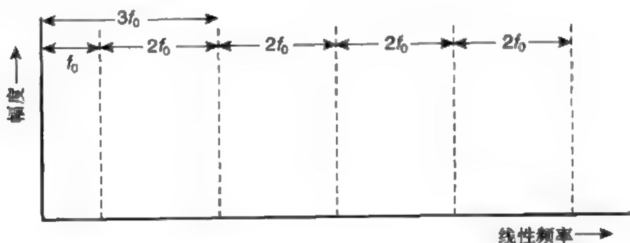


图 3.7 只包含奇次谐波的理想化频谱（对比基频去除前后谐波之间的间隔）

"方法 3" 可以为以下声音找到合适的基频：（1）缺失基频成分的声音（见表 3.3 并忽略基频那一行）；（2）只有奇次谐波的声音（见表 3.3 并忽略偶次谐波那几行）；（3）缺少基频成分只有奇次谐波的声音（见表 3.3 忽略基频和偶次谐波那几行）。每一种情况下的最大公因数都是基频 f_0。这个方法也是解释如何确定非谐和声音高的基本依据，例如铃声、钟声，它们的各频率成分并不是所感知音高频率的精确谐波（不严格满足基频整数倍的关系）。

作为非谐和声音高感知的例子，Schouten 在他最近的一次实验中，合成了一种声音，其频率成分是 1 040Hz、1 240Hz 和 1 440Hz，听音实验得出被感知的音高接近于 207Hz（参考本书 CD 的第 4 条（C））。基于频率成分之间最小频率差的方法（"方法 2"），这些频率成分的基频就是 200Hz。表 3.4 所示为根据 "方法 3" 得到的结果（找出这 3 个频率成分的最大公因数），其中最大除数为 10。Schouten 的目的是利用图表找到一组最接近的数值，那么以这个数值为基频的谐波成分与这 3 个频率成分最接近，然后将它们取平均估算出基频 f_0 的值。按照这种方法，如表 3.4 所示，表中 208Hz、207Hz 和 206Hz 这 3 个数值最接近，将 1 040Hz 视为第 5 次谐波、1 240Hz 为第 6 次谐波、1 440Hz 为第 7 次谐波，它们的基频分别是 208Hz、207Hz 和 206Hz，这些频率的平均值是 207Hz。Schouten 将这种情况下感知的音高称为 "残留音高"，有时也称为 "虚拟音高"。

表 3.4　找到 3 个频率成分 1 040Hz、1 240Hz 和 1 440Hz 的最大公因数以解释约为 207Hz 的实际音高感知

频率成分 (Hz)	÷2 (Hz)	÷3 (Hz)	÷4 (Hz)	÷5 (Hz)	÷6 (Hz)	÷7 (Hz)	÷8 (Hz)	÷9 (Hz)	÷10 (Hz)
1 040	520.0	346.7	260.0	208.0	173.3	148.6	130.0	115.6	104.0
1 240	620.0	413.3	310.0	248.0	206.7	177.1	155.0	137.8	124.0
1 440	720.0	480.0	360.0	288.0	240.0	205.7	180.0	160.0	144.0

最后顺便提一下，我们注意到 1 040Hz、1 240Hz 和 1 440Hz 这些频率成分事实上确实存在一个频率为 40Hz 的基频，它们分别是它的第 26 次、第 31 次和第 36 次谐波，如果表格继续往右延伸 40Hz 将会显现出来。尽管如此，听觉系统似乎更倾向于找到使之成为相邻谐波成分的那个基频 f_0。

3.2.2　部位学说存在的问题

部位学说为解释听觉系统从一组频率中分析出其基频成分提供了理论依据，然而它还存在一系列问题，因为它不能解释以下几点：

（1）人类感知音高的极高精度；

（2）频率成分不能被听觉部位机理分辨的声音的音高感知；

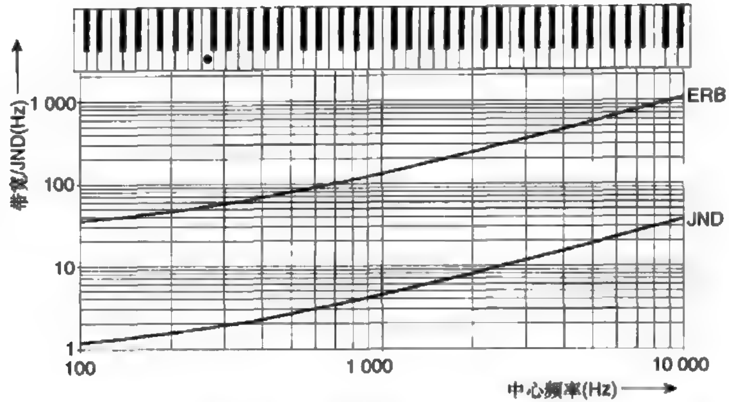
（3）具有连续（非谐波）频谱的声音的音高感知；

（4）基频低于 50Hz 声音的音高感知。

上述几点将逐一进行讨论。

在心理声学中，辨别除了一个参量不同（例如基频、强度、持续时间等）其他参量均相同的两个非常相似的声音的能力可以用“差别阈限”（“difference limen”，缩写为 DL）或“最小可觉差”（“just noticeable difference”，缩写为 JND）来衡量。本书更倾向于使用 JND。人类音高感知的最小可觉差如图 3.8 所示，图中同时表示了临界频带曲线。图中的 JND 是由 Zwicker 等人（1957）实验得出的。在实验中使用正弦波（稳态信号）作为激励信号，并且使声音强度和声音的持续时间保持不变。实验结果表明，JND 在人耳听觉范围内接近临界频带的 1/30。从音乐角度来说，这比较接近于半音程的 1/12。因此，音高的 JND 比听觉滤波器的带宽（临界频带）小得多。

图 3.8　音高感知的最小可觉差（JND）和等效矩形带宽（ERB）



假如相关的听觉滤波器的临界频带足够窄，可以把相邻的谐波成分排除在外，那么听觉的部位机理可以分解出该谐波成分。事实证明，不管声音的基频是多少，它的前 5 ~ 7 次谐波一般都能通过部位分析机理分辨出来。这一点可以通过下面的例子进行说明，并参考表 3.5。

表 3.5　部位机理对基频 f_0 为 110Hz 的前 10 个谐波的分辨能力

谐波频率(Hz)	滤波器临界带宽(Hz)	中间频率(Hz)	临界带宽(Hz)	是否能被分辨
110	36.57	165.0	42.51	是
220	48.45	275.0	54.38	是
330	60.32	385.0	66.26	是
440	72.19	495.0	78.13	是
550	84.07	605.0	90.00	是
660	95.94	715.0	101.8	是
770	107.8	825.0	113.7	否
880	119.6	935.0	125.6	否
990	131.5	1 045	137.5	否
1 100	143.4	1 155	149.3	否

考虑到声音是由多次谐波组成的（f_0、$2f_0$、$3f_0$、$4f_0$、$5f_0$ 等），它们的基频 f_0 是 110Hz，该基频的前 10 个谐波频率列举在表格的最左侧。后面一列是利用式（2.6）以谐波频率为中心频率计算的临界带宽，由此可见临界带宽随着滤波器中心频率的增大而增大（见图 3.8）。基底膜的频谱分析功能相当于一个滤波器组，当临界带宽超过相邻谐波的间距（当所有谐波都存在时，这个间距就是 f_0）时，就不能通过部位分析机理来分辨谐波。在表中可以看到第 7 次谐波 770Hz 的临界带宽小于基频 f_0。这个滤波器能够分辨第 7 次谐波，因为它的中心频率为 770Hz，其带宽的一半扩展到 770Hz 之上，

另一半延伸到 770Hz 之下。为了判断相邻的谐波频率能否被分辨，需要考虑中心频率位于两个相邻谐波中间的听觉滤波器（其中心频率和临界带宽在表 3.5 中列出）。第 7 和第 8 次谐波中间的滤波器具有 113.7Hz 的临界带宽，它大于基频 110Hz，因此听觉不能分辨第 7 和第 8 次谐波。在上述例子中，基底膜不能分辨第 6 次谐波以上的频率成分，因为在第 6 次谐波刺激部位之上的位置，至少有两个相邻的谐波会在基底膜的同一位置产生响应。附录 1 给出了由式（2.6）推出的已知临界带宽求中心频率 f_0 的公式。

例3.4 试进一步证明对于表3.5 中基频为 110Hz 的各次谐波成分，听觉可以分辨第6次谐波，但不能分辨第 7 次谐波。

由附录 1 的式（A1.2）可知：

$$f_c(\text{kHz})=\left|\frac{\left(\frac{ERB}{24.7}\right)-1}{4.37}\right|$$

在临界带宽（ERB）等于 110.0Hz 时，其中心频率（f_c）（请注意该公式中中心频率的单位是 kHz）为：

$$f_c(\text{kHz})=\left|\frac{\left(\frac{110}{24.7}\right)-1}{4.37}\right|=\left|\frac{4.453-1}{4.37}\right|=\left|\frac{3.453}{4.37}\right|=0.79\text{ kHz} \text{ 或 } 790\text{Hz}$$

临界带宽（ERB）等于 110.0Hz 的滤波器的中心频率是 790Hz。因为临界带宽随着中心频率的增大而增大，所有中心频率大于 790Hz 的滤波器的带宽都将大于 110Hz（见图 3.8）。790Hz 小于第 7 次谐波（770Hz）和第 8 次谐波（880Hz）之间的中心频率，因此第 7 次谐波（770Hz）和第 8 次谐波（880Hz）之间的听觉滤波器的带宽大于 110Hz，因而这两个相邻的谐波不能被分辨。所以，听觉可以分辨第 6 次以及第 6 次以下的谐波，但不能分辨第 7 次以及第 7 次以上的谐波。

观察图 3.8 中临界频带和中心频率的关系，我们可以总结出一个一般性规律，即不论 f_0 为多少，听觉不能分辨高于约第 5 次到第 7 次的谐波，原因表述如下。我们可以在图中找到临界频带大于声音基频 f_0 的中心频率，大于这个中心频率的谐波则不能被听觉分辨。这个中心频率可以通过作图的方法得到。沿着 y 轴上的基频位置做一条水平线，通过这条线与临界频带曲线的交点可以在 x 轴上找出相应的中心频率。只有低于这个频率的谐波可以被分辨，高于这个频率的谐波不能被分辨。可以通过选择几个不同的 f_0 对上述一般规律进行验证。这个规律对理解音高感知

以及心理声学其他方面的问题是至关重要的。

一些具有非谐波频谱的声音也可以有音高感，这种说法相悖于前面对音高感知的总体性描述的第二部分："听觉不对时间上非周期性变化的声压波形产生音高感知"。例如，听单词 sea 中 "ss" 和单词 shell 中 "sh"（请尝试朗读这两个单词）的发音，判断哪一个音高较高。大多数听音者会迅速地回答 "ss" 的音高比 "sh" 的高。这两个单词的频谱都是连续的，如图 3.9 所示。请注意，"sh" 频谱能量较多地集中在较低频率处，并且在约 2.5kHz 出现一个峰值，而 "ss" 的峰值出现在 5kHz 附近。这些声音频谱能量 "重心" 的高低是听觉判断这类噪声性声音音调高低的依据。但是听觉对非周期性变化的声音的音高感知能力比周期性变化的声音差得多。这种音高感知现象对理解音乐中非周期性声音的音高感知非常重要，例如，判断打击乐器组中某些乐器产生的声音的音高（参考本书 CD 的第 5 条）。然而绝大部分乐器产生的声压波形是周期性变化的。

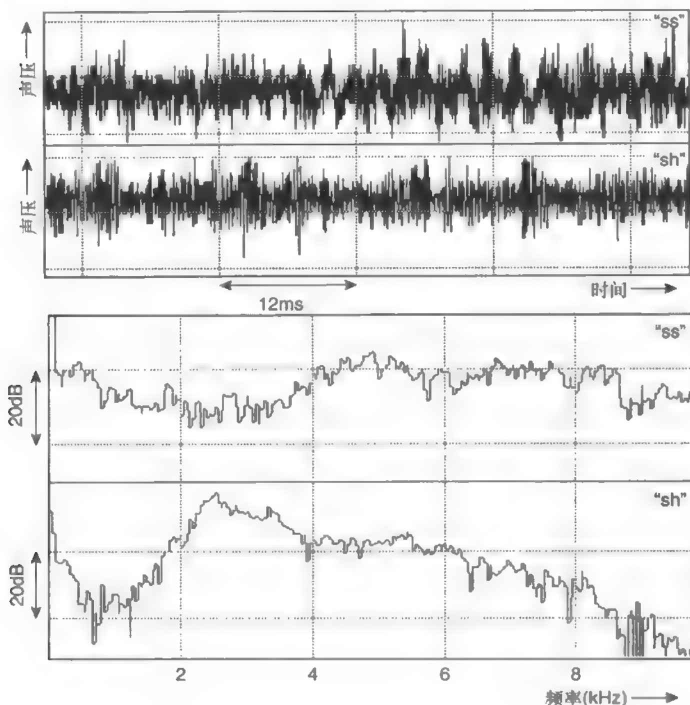


图 3.9　英文单词 sea 中 "ss" 和 shell 中 "sh" 的波形和频谱

最后要说明的是，只包含低于 50Hz 频率成分的声音的音高感知无法用部位学说来解释。因为在这个频率范围，基底膜的振动模式不随频率的改变而改变。这类声音非常少有，但是可以用电子合成器产生。人类最低可听频率为 20Hz，一个基频为 20Hz 的声音的谐波为 40Hz、60Hz 等，因而只有前两个谐波在基底膜上的响应不发生变化，高于 50Hz 的谐波则可以用一般的部位机理分析出来。人类能够对 20Hz ~ 50Hz 的正弦波产生不同的音高感，但是部位机理不能解释这个现象。

还有一些关键问题部位学说理论不能解释，因此下面将要介绍音高感知的理论，它可以和部位学说一起解释部分这样的问题。

3.2.3　音高感知的时间论

音高感知的时间论是基于这个实事，即具有强烈音高感的声音波形是重复的或周期性的（见表 3.2）。图 3.1 列举了 4 种不同乐器演奏 A4 的声压波形。周期性声源的基频 f_0 可以通过测量波形的周期，并用式（3.1）计算得到。

音高感知的时间论依赖于柯蒂氏器（见图 2.3）的神经放电时机，它在基底膜发生振动时就会发出神经电脉冲。部位学说的建立是以输入声波的不同频率成分刺激基底膜不同部位为基础的，而时间论的关键是它详细说明了实际声波沿着基底膜的整个长度刺激不同部位后的具体性质。基底膜可以用一组带通滤波器模拟，这些滤波器的带宽和中心频率的关系如图 3.8 所示。

图 3.10 所示为一组电子带通滤波器的输出波形，这是 Howard 等人（1995）用一些晶片机实现的。输入信号是小提琴演奏的 C4 音，滤波器的带宽是根据式（2.6）的临界频带计算公式得出的。C4 的标准基频为 261.6Hz（见图 3.21），滤波器输出频率刚刚大于 200Hz 的正弦波频率就是基频 f_0，也就是图中最小的那个中心频率。这是因为分析滤波器可以分辨出基频 f_0。任何复杂的周期性波形的每个谐波成分都是正弦波（见第 1 章）。部位学说认为基底膜可以分辨前 6 次谐波（见表 3.5 中的相关计算）。从图 3.10 中可以看出第 2 次（大约为 520Hz）、第 3 次（大约为 780Hz）、第 4 次（大约为 1 040Hz）和第 5 次（大约为 1 300Hz）谐波均能被分辨

出来，而且它们的波形都是正弦波。其中一些正弦波的幅度变化非常明显，尤其是第 4 次和第 5 次谐波能够体现乐器声压输出的动态特性。第 6 次谐波（大约为 1 560Hz）的幅度变化更加显著，但是它的周期性与其他有所不同。

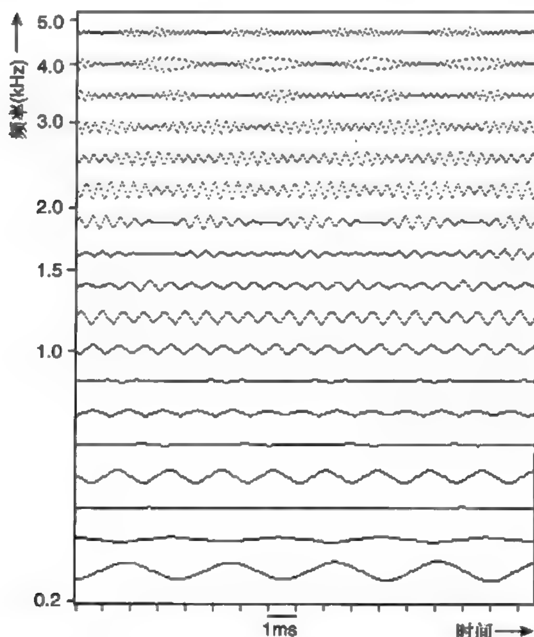


图 3.10　基于晶片机的人类听觉模型的输出说明了基底膜不同部位的振动性质（输入小提琴 C4 音）

在这个例子中，滤波器输出的第 6 次谐波以上的波形均不为正弦波，因为这些谐波不能被分辨。两个或两个以上的谐波合成的信号经滤波器输出后，其波形不为正弦波，如图 3.10 所示。两个频率比较接近的谐波分量合成在一起后，将形成一个幅度有规律变化的“拍音”波形，拍音的频率等于两个分量的频率差。因此如果这两个分量是相邻谐波，拍音的频率就等于基频 f_0。拍音的周期还是（$1/f_0$），这可以从图中对比 1.5kHz 以上的拍音周期和输出基频为 f_0 的正弦波的周期看出。因此当输入包含相邻谐波成分时，滤波器中心频率大于第 6 次谐波的输出波形的周期是（$1/f_0$）。

促使柯蒂氏器进行神经放电的所有滤波器输出波形的周期，是形成音高感知时间论的基础。基底膜上所有部位的神经纤维都可以放电，它们是以这样的方式放电的：某一特定的神经纤维只能在刺激波形周

期的特定相位或特定时间进行放电，这个处理过程称为“锁相（phase-locking）”。尽管神经放电的过程在每个周期都是锁相的，根据观察可知，没有一根神经纤维能以大于 300Hz 的频率进行连续放电。这就是说，神经不一定在每个周期都放电，放电的周期往往是随机的。这种不连续放电的现象是 Pickles（1982）提出的，他认为“也许会少到平均每 100 个周期只放电 1 次”。因为存在锁相，所以每个神经放电的时间间隔都是刺激波形周期的整数倍（1，2，3，4…），并且在基底膜的每个部位都有一定数目的神经纤维在放电。Wever（1949）提出了“群射”理论（“volley firing” principle），即一组神经纤维同时工作，每一根神经纤维以不同的时间周期放电，由此可产生大于 300Hz 的放电频率。还有很多关于这方面较为全面的探讨，这已超出本书讨论的范围，有兴趣的读者可以查阅以下作者的文献资料：Pickles（1982）、Moore（1982，1986）以及 Roederer（1975）。下面将以锁相原理为基础展开进一步的讨论。

在小提琴演奏 C4 时，基底膜不同部位上的神经放电之间的最短时间间隔（刺激波形的一个周期）可以根据图 3.10 推断出，它等于分析滤波器输出波形的周期。对于那些对前 6 次谐波发生响应的部位，其神经放电的最小时间间隔等于该次谐波的周期，对于前 6 次以外的谐波激发的部位，神经放电的最小时间间隔等于输入波形的周期（即 $1/f_0$）。

神经放电可能的时间间隔在图 3.11 中表示出来，该图清晰直观地说明了神经放电的现象。我们可以看出，神经放电对刺激波形而言具有锁相作用，但并非每个周期都有神经放电的动作。图中表示了理想化展开基底膜的部位与各输入谐波频率（最高达到第 16 次谐波）的对应关系，但是并没有明确说明输入声音基频 f_0 的大小。这张图是以如下假设为基础的，即听觉可以分别对第 7 次和第 7 次以下的谐波进行分析。这张图主要表明了，神经放电的间隔以锁相原理为基础，神经放电可能不会在每个周期都出现，用竖线的长度表示在基底膜上特定位置和特定时间间隔出现神经放电的概率大小。由图可知，时间间隔较短的放电概率较大。该图与神经放电条形图（histogram of firings）很相似。神经放电条形图有时也称为“放电间隔（inter-spike interval）”条形图，图中一根“长针”

表示一次神经放电。

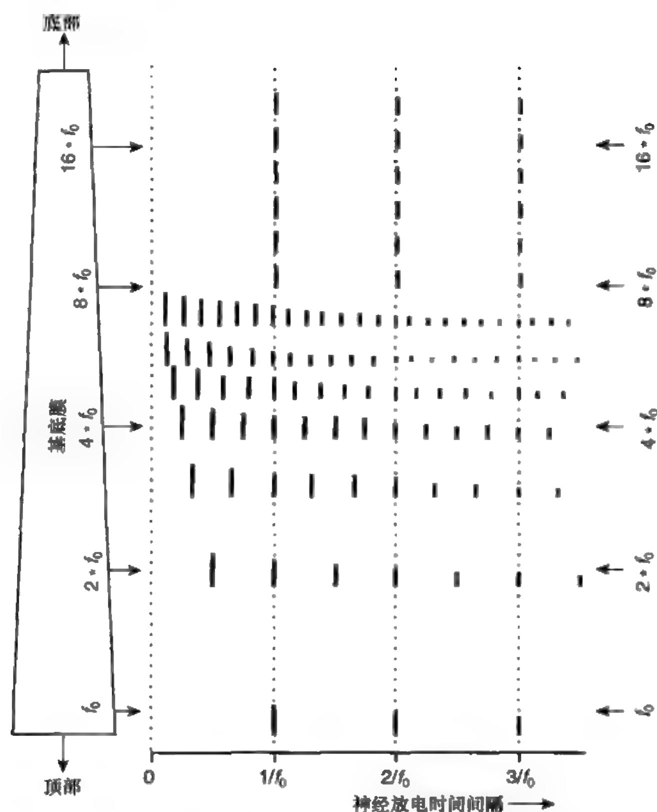


图 3.11 基底膜对输入声音的前 16 个谐波可能的神经放电时间间隔

因此，图中基底膜相应部位在受到基频 f_0 刺激后的神经放电时间间隔可能是（$1/f_0$）、（$2/f_0$）和（$3/f_0$），较大时间间隔的神经放电次数较少（出现的概率较小）。对于第 2 次谐波刺激的部位，可能的放电时间间隔为 $[1/(2f_0)]$、$[2/(2f_0)]$ 或 $(1/f_0)$、$[3/(2f_0)]$、$[4/(2f_0)]$ 或 $(2/f_0)$ 等。这个规律适用于第一次谐波到第 7 次谐波。对于第 7 次以上的谐波，刺激波形就像拍音一样，而且它们的周期是（$1/f_0$）。因此如图所示神经放电的可能时间间隔为（$1/f_0$）、（$2/f_0$）和（$3/f_0$）。从图中可以想象，如果所有的这些“放电间隔”条形图沿垂直方向叠加到一起（对每个放电时间间隔而言），那么最大放电概率将出现在时间间隔等于频率 f_0 的周期上。当考虑到更高的谐波成分可能存在，并且以和第 8 次谐波相同的规律进行神经放电时，这种现象会进一步得到加强。需要注意的是，基底膜上能够分辨谐波的

每个部位在每个基频周期都会出现神经放电，这正是神经纤维并不是在每个周期都放电的直接结果。这是音高感知的时间论能够解释听觉音高感知现象的基本出发点。

3.2.4　时间论存在的问题

时间论为理解听觉如何通过分析基底膜上各部位的神经放电时机找到基本周期提供了理论基础。然而，不是仅仅依靠时间论就能解释人类所有音高感知的能力。听觉感知基频大于 5kHz 的声音音高的能力不能用时间论来解释，因为当基频大于 5kHz 时听觉就会失去锁相功能。基频大于 5kHz 的音高感知能力全部归因于部位学说。鉴于人类存在听觉上限频率——年轻人是 20kHz，20 岁以上人群则下降为 16kHz，因此基频大于 5kHz 的声音只能为听觉系统提供两个可分析的谐波（f_0 和 $2f_0$）。实际上我们都知道，人类感知基频大于 5kHz 的声音的能力非常薄弱，许多音乐家也很难准确判断出这个频率范围的音程关系。Moore（1982）指出这个现象与短笛的上限频率约为 4.5kHz 比较吻合。对于大型管风琴，一些音栓控制的音管的基频甚至超过 8kHz，但是这些音管一般只用来配合其他音栓一起发声（见第 5.4 节）。

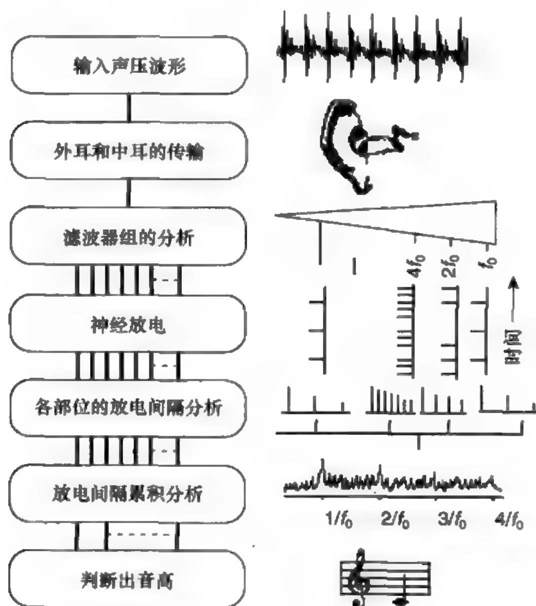
3.2.5　音高感知的当代理论

心理声学研究在历史上倾向于用部位学说和时间论来解释人类的音高感知能力，而我们知道没有一个理论能够单独解释所有音高感知现象。事实上，部位学说解释了基底膜上每个部位受到刺激后产生神经放电机理，因此神经中枢和一部分与听觉处理有关的大脑不仅得到了基底膜受到刺激部位的信息（频率分析），同时得到了有关刺激特性的信息（时间分析）。因此没有一个理论能够单独完整地解释所有音高感知的能力，因为不管是部位分析还是时间分析，它们的输出在听觉神经的传输和处理过程中都用到了另一个方面的信息。

图 3.12 所示为复音的音高感知模型，它是 Moore（1982）提出的，它提取了部位学说和时间论两者的优点。声压波形经过外耳和中耳传输后

被改变（见第 2 章），然后部位机理将分析频谱成分，它的作用可以用一个滤波器组来等效。基底膜上与特定频率成分相对应的部位发生振动后转化为神经电脉冲，放电方式以锁相原理为基础，但不一定每个周期都放电。后者对应于图中右侧的图示。神经放电发生在特定部位的事实为音高感知的部位学说奠定了基础。神经冲动的时间间隔被听觉系统分析后，得到一些共同的时间间隔。这些时间间隔接近于基频的周期及其倍数，但主要集中在（$1/f_0$）上，这是音高感知时间论的基础。听觉根据上述分析的结果判断声音的音高。

图 3.12 Moore（1982）提出的人类听觉音高感知模型



3.2.6 影响音高感知的其他因素

音高的感知主要受到基频 f_0 的影响，这就是为什么音符的音高通常与 f_0 有直接的关系，如规定 A4 以基频 440Hz 作为音高参考标准。基频大小的测评成为音高感知的部位学说和时间论的基础。从某个音程中音高的变化可以确定相关音符的基频是否处在合适的比例关系，以便为音高感知（主观的）进而音程感知提供主要的声学（客观的）依据。另一方面，如果保持基频不变，当改变声音的强度和持续时间

时，音高感觉也会发生变化。除了改变基频 f_0 会影响音高的感知外，到目前为止，上述几点是影响音高感知的其他因素，这些影响是十分微妙的。

影响音高感知的其他因素可以总结如下。当保持基频 f_0 不变时，如果正弦波的强度在 40dBSPL ~ 90dBSPL 变化，所有除 1 ~ 2kHz 以外的基频的音高感知会随之变化。当基频大于 2kHz 时，音高感知随着强度的增大而变高；当基频小于 1kHz 时，音高感知随着强度的增大而变低。这种现象可以体现在图 3.13 中。音高的最小可觉差 JND 以 60dBSPL 时的值为参考值，使不同频率和强度的正弦波的音高变化可以相互比较。图 3.13 是针对正弦波的测量结果，而正弦波在音乐中很少单独出现，尽管电子合成器可以轻易地产生这种声音。复音音高受强度影响的研究较为少见，Rossing(1989) 提出复音声强在 65dBSPL ~ 95dBSPL 变化时，音高变化大约为 17 音分 (0.17 个半音)。Rossing 针对它给音乐带来的影响提出了两个建议：(1) 他引用 Parkin(1974) 的研究结果，指出在一个混响效果显著的房间里听管风琴响度很大的结尾和弦时，这种音高变化的现象较为明显，即随着声压级逐渐减小音高反而变高；(2) 他指出在基频 f_0 保持不变时，随着波形幅度变化速率的变化，声音的音高会随之偏移，这种现象“在处理打击乐器时应该考虑到”。

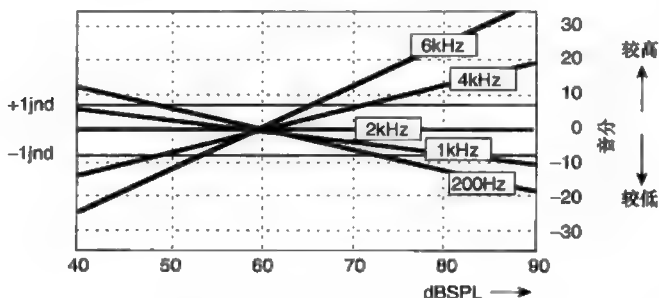


图 3.13 保持基频不变时正弦波的音高随强度的变化特性 (Rossing 1989)

声音的持续时间对音高感知的影响不是一个简单关系，但是这种关系可以用图 3.14 表示出来。图中给出了基频 f_0 为特定值时能在听觉产生音高感的最小周期数目。较短持续时间的声音可能被感知具有一定音高，而不会感觉不到音高；但是当声音持续时间一旦低于图 3.14 中最小周期

数时，听音者判断音高的准确性会变差。

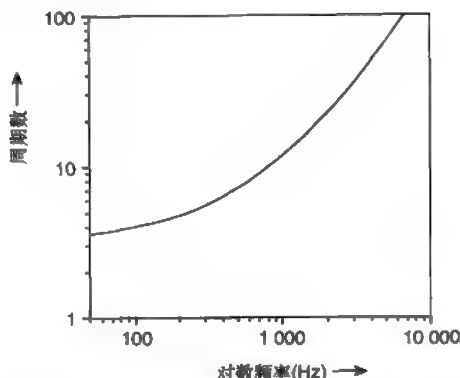


图 3.14 听觉能够感知特定频率音高所需的最小周期数（Rossing 1989）

在本节的最后需要简要说明的是音高感知的“重复音高（repetition pitch）”现象，尤其在现在电子合成器和演播室技术发展日新月异的年代，重放这种声音变得十分容易（参考本书 CD 的第 6 条）。如果信号源为非周期的噪声性信号，如瀑布声。单词 see、shoe、fee 中的辅音或者噪声发生器，将这些信号加延时后叠加到原始信号上，然后重放给听音者听，大多数听音者会感觉到“重复音高”的存在，但并非所有听音者都能感觉到。如果改变延时的大小，音高的感知也会发生变化，感知的音高就等于以延时为周期的基频 f_0 的音高。能够产生“重复音高”现象的延时大约在 1ms ~ 10ms，这个值因人而异，对应于基频为 100Hz ~ 1 000Hz 的音高。借助于现代电子设备，用这种效果对声音进行变调处理是完全可能的。

3.3 乐音的听音

不同文化背景的音乐可能在很多方面有较大的区别，包括音调、节奏、编曲、动态范围、基本旋律和音乐中和声的应用等。音乐品味总是随着时间的推移而不断变化，一个作曲家的尝试在若干年后可能成为一种既定的传统。本节将从心理声学角度讨论和弦的感知和不同调律系统的发展，作为本章前面对西方音乐的旋律与和声发展的声学讨论的补充。

3.3.1　谐波以及西方和声的发展

听辨和声是音乐欣赏的基础，并且和声的基础在于对和弦的运用。和弦是由两个或两个以上同时发声的音符构成的，和弦可以用组成和弦的音符之间的音程来描述。

研究两个同时发声的音符引起的心理感知是理解和弦心理声学的基础。图 3.15 所示为相对于中央 C 的一组完整且常见的双音音程及其名称，其名称如音符下方所示。如果把一些音程在现代键盘上弹奏出来，可以看出每个增加的音程和相应减少的音程具有相同的听感，例如增一度和小二度、增四度和减五度、增五度和小六度以及大七度和减八度。然而它们的五线谱标记方法却是不同的，如果所用的调律系统不同，那么这些异符同音的音程也可能具有不同的听感。

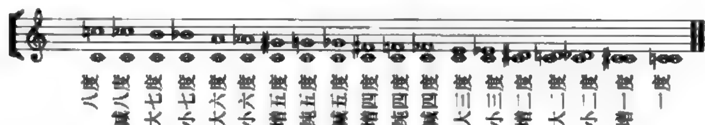


图 3.15　八度以内以 C4 为根音的所有双音音程

从自然谐波系列中相邻谐波的音程随谐波次数的增加而逐渐减少的趋势，可以看出和声在西方音乐的发展。图 3.3 所示为自然谐波系列中前十次谐波的音程关系。随着谐波次数的增加，相邻谐波的音程是逐渐减小的，因为音程取决于两个谐波基频 f_0 的比例关系（如 2 ： 1>3 ： 2>4 ： 3>5 ： 4>6 ： 5 等）。

西方最早时期的复调音乐，称为“复音乐曲”，主要利用了八度、纯五度和它的转位纯四度的音程关系。这些音程分别由自然谐波系列的第 1 次和第 2 次、第 2 次和第 3 次、第 3 次和第 4 次构成（见图 3.3）。后来出现并慢慢被接受的有大、小三度，其音程是由第 4 次和第 5 次、第 5 次和第 6 次自然谐波构成的，以及它们的转位大、小六度，分别由第 5 次和第 8 次、第 3 次和第 5 次谐波构成。大三和弦是由一个大三度和一个小三度共同构成，小三和弦是由一个小三度和一个大三度共同构成。大小三和弦成为西方调式和声的主要组成部分。此后小七度的音程和它的转位大二度，对应于第 4 次和

第7次谐波以及第7次和第8次谐波构成的音程，也开始被广泛采纳。20世纪的作曲家开始探索用所有的音程创作音乐。第8次和第9次谐波、第9次和第10次、第11次以上的相邻谐波的音程是接近半音和微分音（音程小于半音）的，高于第16次谐波的谐波间隔是由微分音衡量的。

3.3.2 谐和与不谐和

西方和声按照一定的模式发展，在这种模式下音程作为音乐发展的中心逐渐提升了自然谐波系列的作用。这个变化部分是由于人们越来越接受被视为在音乐上"谐和的"或听起来"愉悦的"音程，这些音程与"不和谐的"或听起来"不愉悦的"是相对的。"谐和"与"不谐和"背后的心理声学基础与临界带宽有关，它可以衡量音程的"谐和"（或"不谐和"）程度。

图2.6表示当两个正弦波的频率差从0增大至一个临界带宽时听感的变化情况。当频率差超过一个临界频带时，听音者的听感从"不平稳"变为"平稳"。此外，在临界带宽范围内随着频率差逐渐增大，听音者的听感从"不平稳混合在一起"变为"不平稳可分离"。图3.16是Plomp和evelt（1965）的实验结果。这个实验确定了当两个正弦波频率差为多大时，具有"谐和"或"不谐和"的听觉感受。实验要求没有受过音乐训练的受试者指明这两个正弦波组成的声音的谐和性和愉悦程度。（由于音乐家对音程的谐和程度事先有所了解，所以该实验中没有使用他们作为受试者。）实验结果如图中所示，曲线是连续的，不存在谐和程度突出的特殊音程，并且大于小三度的所有频率比的音程都是谐和的。由此可以得出以下的结论：

（1）当两个纯音的频率相等（同音）时，这两个纯音被感知为"完全谐和音程"；

（2）当两个纯音的频率差大于一个临界带宽时，这两个纯音被感知为谐和音程；

（3）当两个纯音的频率差处于临界带宽5%～50%时，这两个纯音被感知为不谐和音程；

（4）当两个纯音的频率差是临界带宽的 1/4 时，这两个纯音的不谐和性达到最大。

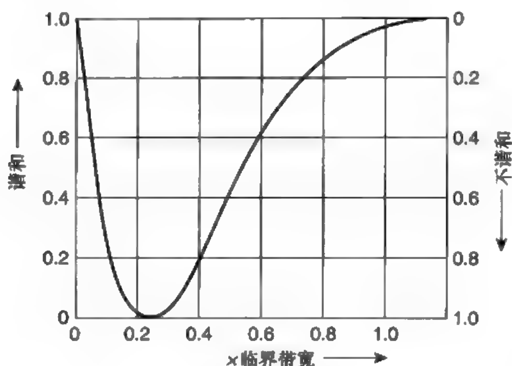


图 3.16 两个纯音信号谐和与不谐和的感知规律（Plomp 和 evelt 1965）

由于几乎没有一种乐器可以产生单一正弦波的波形，因此可以将有关纯音谐和与不谐和的研究结果扩展到听觉感知乐器同时发出多个周期性复音的情况。对于和弦中的每个音符，如果每个音符单独演奏出来，听觉能够分辨所有不高于第 7 次的谐波，所有的这些谐波将有助于感知声音的谐和性或不谐和性，其判断取决于该谐波频率与和弦中另一个音符的谐波频率相近程度。前 7 次谐波对谐和性的贡献可以根据图 3.16 进行评价，和弦总体的谐和（不谐和）性是以每个谐波对的谐和（不谐和）性为基础。

3.3.3 音程的听辨

以心理声学为基础，音程可以按照谐和程度逐渐减小的顺序来排列。为了确定两个复音的谐和程度，假设两个复音分别包含所有的谐波成分，首先找出低于较低音符的第 7 次（包含第 7 次）谐波频率的所有谐波频率，然后找出每对最相近谐波的中间频率及其临界频带，通过判断这两个谐波是否在这个临界带宽的 5% ~ 50% 范围内，可以确定这两个谐波对整体不和谐性感知的贡献。如果冠音的谐波位于根音的相邻谐波中间，由于临界带宽随频率的增大而增大，所以应该选取频率较高的谐波对考虑其谐和性，可以从表 3.6 看出这一做法。（这种分析方法与表 3.5 相似）

表 3.6　纯五度双音和弦的谐和与不谐和程度（根音基频为 220Hz，包含所有谐波成分）

纯五度音程（3 ： 2）（根音基频为 220Hz）						
根音的前 7 次谐波 (Hz)	冠音的谐波 (Hz)	频率差 (Hz)	中间频率 (Hz)	临界带宽 (Hz)	半临界带宽 (Hz)	完全谐和音程（C） 谐和音程（c） 不谐和音程（d） 极不谐和音程（D）
220						
440	330	110	385	66.3	33	c
660	660	0	同音	–	–	C
880	990	110	1 045	138	69	d
1 100						
1 320	1 320	0	同音	–	–	C
1 540	1 650	110	1 595	197	99	d

例如，表 3.6 计算了基频 f_0 为纯五度的两个音符（基频比为 3 ： 2）及其谐和性的判断结果，根音的基频为 220Hz。根音的各次谐波与最接近的冠音谐波的频率差在表中第 3 列列举出来（如果某个谐波处在两个谐波的中间位置，则取较高频率的两个谐波配对），这两个谐波的中间频率值表示在第 4 列中，相应中心频率的临界带宽表示在第 5 列中。每个谐波对对音程谐和性的贡献表示在最右边一列，分别说明如下：

（1）如果谐波对的两个谐波音高相同（频率相等），它们是“完全谐和音程”，用“C”表示（因为它们的频率差小于临界带宽的 5%）；

（2）如果谐波对的频率差大于中间频率的临界带宽，它们是“谐和音程”，用“c”表示（也就是第 3 列的数值大于第 5 列的数值）；

（3）如果谐波对的频率差小于中间频率的临界带宽的一半，它们是“极不谐和音程”，用“D”表示（也就是第 3 列的数值小于第 6 列的数值）；

（4）如果谐波对的频率差小于中间频率的临界带宽，但是大于带宽的一半，它们是“不谐和音程”，用“d”表示（也就是第 3 列的数值小于第 5 列的数值，但是大于第 6 列的数值）。

音程对不谐和性的贡献与音程出现在自然谐波系列中的位置有关。音程出现在谐波系列中的频率位置越高，它对不谐和性的贡献越大。在

同音的两个音符构成的音程中，两个音符的所有谐波都是谐和的，每个谐波对的贡献均为“C”。对于倍频程，冠音的所有谐波与根音的谐波均同音，因此倍频程的贡献为“C”。表 3.6 ~ 表 3.10 分别列出了纯五度（3 ∶ 2）、纯四度（4 ∶ 3）、大三度（5 ∶ 4）、小三度（6 ∶ 5）、大全音（9 ∶ 8）对音程的谐和性和不谐和性的贡献情况。每个不同音程和弦的不谐和性都在表中最后一列表示出来。谐波的不谐和性以 C、c、d、D 的顺序依次增加，从图表中可以看出，随着谐波次数的增加以及音程的减小，不谐和性会增加。

表 3.7　纯四度双音和弦的谐和与不谐和程度（根音基频为 220Hz，包含所有谐波成分）

纯四度音程（4 : 3）（根音基频为 220Hz）						
根音的前 7 次谐波 (Hz)	冠音的谐波 (Hz)	频率差 (Hz)	中间频率 (Hz)	临界带宽 (Hz)	半临界带宽 (Hz)	完全谐和音程 (C) 谐和音程 (c) 不谐和音程 (d) 极不谐和音程 (D)
220	293	73.0	330	60	30	c
440						
660	586	73.3	623	92	46	d
880	879	同音	–	–	–	C
1 100	1 172	73.3	1 170	151	76	D
1 320						
1 540	1 465	73.3	1 500	187	93	D

表 3.8　三度双音和弦的谐和与不谐和程度（根音基频为 220Hz，包含所有谐波成分）

大三度音程（5 : 4）（根音基频为 220Hz）						
根音的前 7 次谐波 (Hz)	冠音的谐波 (Hz)	频率差 (Hz)	中间频率 (Hz)	临界带宽 (Hz)	半临界带宽 (Hz)	完全谐和音程 (C) 谐和音程 (c) 不谐和音程 (d) 极不谐和音程 (D)
220	275	55.0	248	52	26	c
440						
660	550	110	605	90	45	c
880	825	55.0	853	117	58	D
1 100	1 100	同音	–	–	–	C
1 320	1 375	55.0	1 350	170	85	D
1 540	1 650	110	600	197	89	d

表 3.9 小三度双音和弦的谐和与不谐和程度（根音基频为 220Hz，包含所有谐波成分）

小三度音程（6：5）（根音基频为 220Hz）						
根音的前 7 次谐波（Hz）	冠音的谐波（Hz）	频率差（Hz）	中间频率（Hz）	临界带宽（Hz）	半临界带宽（Hz）	完全谐和音程（C） 谐和音程（c） 不谐和音程（d） 极不谐和音程（D）
220	264.0	44.0	242	51	26	d
440	528.0	88.0	484	77	39	d
660						
880	792.0	82.0	833	115	58	d
1 100	1 056.0	44.0	1 080	141	71	D
1 320	1 320.0	同音	–	–	–	C
1 540	1 584.0	44.0	1 560	193	97	D

表 3.10 大全音双音和弦的谐和与不谐和程度（根音基频为 220Hz，包含所有谐波成分）

大全音音程（9：8）（根音基频为 220Hz）						
根音的前 7 次谐波（Hz）	冠音的谐波（Hz）	频率差（Hz）	中间频率（Hz）	临界带宽（Hz）	半临界带宽（Hz）	完全谐和音程（C） 谐和音程（c） 不谐和音程（d） 极不谐和音程（D）
220	247.5	27.5	234	50	25	d
440	495.0	55.0	477	76	38	d
660	742.5	82.5	701	100	50	d
880						
1 100	990.0	110	1 050	138	69	d
1 320	1 237.5	82.5	1 280	163	82	d
1 540	1 485.0	55.0	1 510	188	94	D

那些同音的谐波次数可以推算出来。例如，在纯四度中，因为两个基频的频率比是（4 ： 3），因此根音的第 4 次谐波与冠音的第 3 次谐波音高相同。对于频率比为（9 ： 8）的大全音，根音的第 9 次谐波与冠音的第 8 次谐波音高相同，但是这两个谐波都不能在各自的音符中被听觉分辨出来。

最后要说明的是，由于临界带宽随中心频率变化而变化（见图 3.8），因此给定音程的和弦的谐和程度会随根音基频的变化而变化。表 3.11 和表 3.12 以大三度为例对这个特性进行了说明，其中根音的基频比表 3.8 中的分别低一个倍频程和两个倍频程，即基频分别为 110Hz 和 55Hz。通过

比较可知，随着音程基频 f_0 的降低，结果为 "D"（极其不和谐音程）的数量在增多。

表 3.11　大三度双音和弦的谐和与不谐和程度（根音基频为 110Hz，包含所有谐波成分）

大三度音程（5：4）（根音基频为 110Hz）						
根音的前 7 次谐波（Hz）	冠音的谐波（Hz）	频率差（Hz）	中心频率（Hz）	临界带宽（Hz）	半临界带宽（Hz）	完全谐和音程（C） 谐和音程（c） 不谐和音程（d） 极不谐和音程（D）
110	137.5	27.5	124	38	19	d
220						
330	275.0	55.0	303	57	29	d
440	412.5	27.5	426	71	36	D
550	550.0	同音	–	–	–	C
660	687.5	27.5	674	97	49	D
770	825.0	55.0	798	111	56	D

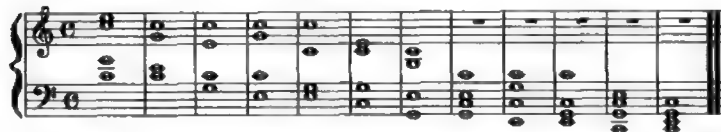
表 3.12　大三度双音和弦的谐和与不谐和程度（根音基频为 55Hz，包含所有谐波成分）

大三度音程（5：4）（根音基频为 55Hz）						
根音的前 7 次谐波（Hz）	冠音的谐波（Hz）	频率差（Hz）	中心频率（Hz）	临界带宽（Hz）	半临界带宽（Hz）	完全谐和音程（C） 谐和音程（c） 不谐和音程（d） 极不谐和音程（D）
55.0	68.75	13.8	61.9	31	16	D
110						
165	137.5	27.5	151	41	21	d
220	206.3	13.8	213	48	24	D
275	275.0	同音	–	–	–	C
330	343.8	13.8	337	61	31	D
385	412.5	27.5	399	68	34	D

除了同音和八度音以外，这种给定音程的不谐和性随音高降低而增强的现象，已经在古典音乐所偏好的和弦音程中得到证实。在写四声部和声如 SATB（soprano 高音、alto 中音、tenor 次中音、bass 低音）赞美诗时，一般要求低音和次中音的音程不小于四度（除非它们处在低音五线谱之上），否则声音会显得"浑浊"或"刺耳"。图 3.17 所示为 C 大调中不同音符间隔构成的四和弦以及它们的转位。在演奏这些和弦时，我们可以听出这些和弦的和声效果。最好能用乐器连续稳定

地弹奏出这些音符，这些乐器可以选择管风琴、乐器组或使用合适的合成声源。为了充分认识到这一点的重要性，最基本的方法是倾听它们的效果，毕竟音乐的心理声学与音乐如何被感知有关，而不仅仅是写在纸上的东西。

图 3.17 C 大调中不同的四和弦（演奏并审听每一个和弦的“浑浊”或“刺耳”程度）



3.4 音律

音阶是西方音乐的基础。现代键盘乐器在每个倍频程上都有 12 个音符，其中每两个邻近的音符间隔为一个半音。所有常用的西方音乐的音阶都包含倍频程，其频率比为（2 ：1）。因此，仅仅考虑一个倍频程内的音阶就够了，因为其他倍频程内的音符频率可以从这个倍频程的频率推导出来。早期的音阶是以自然谐波系列形成的一个或多个音程为基础的（见图 3.3）。

3.4.1 五度相生律（毕达哥拉斯音律）

五度相生律是通过组合纯五度形成的。例如起始音为 C，以纯五度为一个间隔，向上递增 12 次将会形成一个“五度循环”，即 C、G、D、A、E、B、F#、C#、G#、D#、A#、E#、c。这个五度循环的最后一个音符是 c，它与起始音 C 的频率比为（3 ：2）的 12 次方，即：

$$\frac{c}{C}=\left(\frac{3}{2}\right)^{12}=129.746$$

12 个纯五度的音程约等于 7 个倍频程的音程，因为音符（c'）与比它低 7 个倍频程的 C 的频率比为：

$$\frac{c'}{C}=2^{7}=128.0$$

因此，在五度相生律中，12 个纯五度（C 到 c）略高于 7 个倍频程（C 到 c'），此音程差称为“毕达哥拉斯间隔（Pythagorean comma）”，它的频率比为：

$$\frac{c}{c'}=\frac{129.746}{128.0}=1.013\ 64$$

如果五度循环通过递减 12 次而不是递增 12 次来形成，那么递减后的音符将比 7 个倍频程的起始音符低 1.013 643 3。由递减形成的五度循环中的每一个音符与递增形成的每一个音符的音高有微小的差异，图 3.18 表示递增和递减的不同情况以及相对应的音名。例如，音符 #D 与 E♭、#A 与 B♭、B𝄫与 A 的音名不同，但是表示的是同一个音，称为“异符同音”，于是就形成了图 3.15 中的大三度、减四度、大七度和减八度，它们都属于异符同名。通过由起始音符 C 分别构成一个准确纯四度和纯五度可以建立起一个五度相生律音阶（分别与下属音和属音保持这种完美关系）：

$$\frac{F}{C}=\frac{4}{3}$$

$$\frac{G}{C}=\frac{3}{2}$$

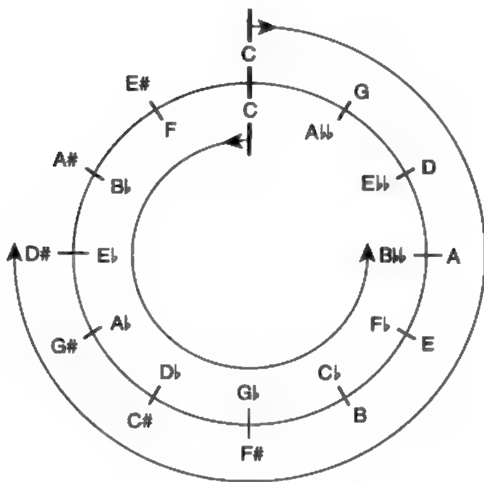


图 3.18　递增 12 个纯五度（外环）或递减 12 个纯五度（内环）构成的五度相生律

音阶中其他音符之间构成的频率比可以通过从 G 向上构成纯五度确

定，必要的时候需要通过降低一个倍频程以保证两个音符在同一个倍频程内。因此音符与起始音 C 构成的频率比分别为：

$$\frac{D}{C}=\frac{3}{2}\times\frac{G}{C}\times\frac{1}{2}=\frac{3}{2}\times\frac{3}{2}\times\frac{1}{2}=\frac{9}{8}$$

$$\frac{A}{C}=\frac{3}{2}\times\frac{D}{C}=\frac{3}{2}\times\frac{9}{8}=\frac{27}{16}$$

$$\frac{E}{C}=\frac{3}{2}\times\frac{A}{C}\times\frac{1}{2}=\frac{3}{2}\times\frac{27}{16}\times\frac{1}{2}=\frac{81}{64}$$

$$\frac{B}{C}=\frac{3}{2}\times\frac{E}{C}=\frac{3}{2}\times\frac{81}{64}=\frac{243}{128}$$

五度相生律大调音阶中的音符与 C 形成的频率比如图 3.19 所示，通过将这些音符与 C 的频率比相除可以得到相邻音符之间的频率比，例如：

$$A\text{ 与 }B\text{ 的频率比}=\frac{243}{128}\div\frac{27}{16}=\frac{243}{128}\times\frac{16}{27}=\frac{9}{8}$$

$$E\text{ 与 }F\text{ 的频率比}=\frac{4}{3}\div\frac{81}{64}=\frac{4}{3}\times\frac{64}{81}=\frac{256}{243}$$

图 3.19 列举了上述大调音阶中相邻音符的频率比，大调音阶包括下列音程：全音、全音、半音、全音、全音、全音、半音，从图中可以看出：

图 3.19 五度相生律大调音阶中各音符与 C 的频率比

音符	C		D		E		F		G		A		B		C
音符与C的频率比	1/1		9/8		81/64		4/3		3/2		27/16		243/128		2/1
音符之间的频率比		9/8		9/8		256/243		9/8		9/8		9/8		256/243	

$$\text{半音的频率比}=\frac{256}{243}$$

$$\text{全音的频率比}=\frac{9}{8}$$

3.4.2 纯律（自然律）

另一种重要的调律方式是通过保持 3 个主要音程关系得到的，即在主音、属音和下属音的基础上形成八度（2 ∶ 1）、纯五度（3 ∶ 2）和大

三度（5 ： 4），属音和下属音的基音分别向上和向下与主音形成一个纯五度。于是就形成了大调音阶中的每一个音符（每一个音符都可以构成这 3 种和弦）。为方便起见，在此选取 C 为起始音可以建立一个大调音阶，如下所述。由主音 C 构成一个大三度（5 ： 4）和纯五度（3 ： 2）可以分别得到音符 E 和 G，即：

$$\frac{E}{C}=\frac{5}{4}$$

$$\frac{G}{C}=\frac{3}{2}$$

B 和 D 分别与主音 G 构成大三度（5 ： 4）和纯五度（3 ： 2），它们与 C 的关系为（计算 D 与 C 的频率比时，将 D 向下降低一个倍频程，以保证与 C 在一个八度以内）：

$$\frac{B}{C}=\frac{5}{4}\times\frac{G}{C}=\frac{5}{4}\times\frac{3}{2}=\frac{15}{8}$$

$$\frac{D}{C}=\frac{3}{2}\times\frac{G}{C}\times\frac{1}{2}=\frac{3}{2}\times\frac{3}{2}\times\frac{1}{2}=\frac{9}{8}$$

A 与 C 分别与下属音 F 向上构成大三度（5 ： 4）和纯五度（3 ： 2）（C 与上方 F 构成了纯四度（4 ： 3），因为纯四度和纯五度相加得到一个倍频程），即：

$$\frac{F}{C}=\frac{4}{3}$$

$$\frac{A}{C}=\frac{5}{4}\times\frac{F}{C}=\frac{5}{4}\times\frac{4}{3}=\frac{20}{12}=\frac{5}{3}$$

在自然大调音阶中 [注 1]，各音符与 C 形成的频率比以及相邻音符的频率比（通过将这些音符与 C 的频率比相除可得相邻音符的频率比）如图 3.20 所示。该图说明自然大调音阶（全音、全音、半音、全音、全音、全音、半音）具有相等的半音程，但是具有两种不同的全音程，音程较大的称为“大全音”，音程较小的称为“小全音”，即：

$$\text{自然音阶半音的频率比}=\frac{16}{15}$$

$$\text{自然音阶大全音的频率比}=\frac{9}{8}$$

$$自然音阶小全音的频率比 = \frac{10}{9}$$

这两个全音和半音的音程就是自然谐波中相邻谐波之间的音程（见图 3.3）。这意味着对于有旋律并且和谐的音乐片段而言，音阶中音符之间应尽可能地保持谐和状态。然而两种不同全音音程的出现意味着这个音阶只能被定在一个基调上，每一个不同的基调都需要重新建立音阶。例如，D 和 A 之间的音程是：

图3.20　C大调自然音阶中各音符与主音C的频率比

音符	C	D	E	F	G	A	B	C
音符与C的频率比	$\frac{1}{1}$	$\frac{9}{8}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{15}{8}$	$\frac{2}{1}$
音符之间的频率比	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{9}{8}$	$\frac{16}{15}$	

$$\frac{A}{D}=\frac{5}{3}\div\frac{9}{8}=\frac{5}{3}\times\frac{8}{9}=\frac{40}{27}$$

这个音程比纯五度音程（3 ∶ 2）略小。

为了满足乐器演奏的实际需要，使其能够在不同基调上弹奏出来，“毕达哥拉斯间隔”需要合理分配在五度循环中以保证音符向上递增 12 个纯五度后，其频率与递增 7 个倍频程的频率完全相等（见图 3.18）。具体做法是将部分纯五度音程减小，同时保持另外一些纯五度音程；或者不同程度地减小所有纯五度的音程；或者甚至是额外地增大一些纯五度的音程和减小纯五度音程使之互相补偿。所以有无数种可能性，但是没有一种能在所有基音上形成准确的自然音阶。对很多调律系统进行实验后的结果是，可以在一些基音上获得接近于自然音阶的三度和五度的音程关系，但是完全破坏了其他基音的音阶特性。Padgham（1986）对调律系统开展了一个更为充分的探讨，他对一些键盘乐器进行了实验，将黑键进行分割（高、低两个方向），产生了异符同音的音符，例如 C# 和 D♭、D# 和 E♭、F# 和 G♭、G# 和 A♭、A# 和 B♭。Padgham 进行实验的乐器包括爱丁堡大学音乐学院的 McClure 管风琴。但是这些调律的探讨并没有在键盘演奏者中得到普及。

［注 1］自然音阶包括大全音和小全音。

3.4.3　十二平均律

由于五度相生律是以五度循环延伸下去的，因此会形成“不平均性”。还有一种可行的方法是以等音程的音符构成音阶，使得这种音程关系对任何一个基音都能得以保持，但是每个音符都与自然音阶存在一定差别。这种调律系统在现代键盘乐器中最为常见，其所有的半音都等于 1/12 倍频程，所以平均律的半音频率比（r）是一个常数，这个常数的 12 次方等于 2，或者表示为：

$$r=\sqrt[12]{2}=1.059\,5$$

半音可以细分为“音分”[注 1]，一个音分等于半音的 1/100。因此，一个音分的频率比（c）为：

$$c=\sqrt[100]{r}=\sqrt[100]{1.059\,5}=1.000\,578$$

在讨论音程或进行关于音高的心理声学实验时，经常会用到音分。附录 2 给出了频率比和音分之间的换算公式。

在十二平均律中，所有半音和全音的频率比是恒等的，因此一段音乐可以用任意音调演奏出来。但是没有一个音程等于自然谐波系列的相邻音程（见图 3.3），因此没有一个音程是完全谐和的[注 2]。然而平均律音阶中的音程仍然可以根据前述准则判断其谐和与不谐和程度。尽管在十二平均律中音符的谐波对不会完全相等，而纯音程的谐波频率可能相同，如表 3.6 ～表 3.12 所示，但是谐波的频率差在临界带宽的 5% 范围内仍然是谐和的。在十二平均律中，和弦的谐波之间会引起拍音现象（见图 2.6），但在自然律中不容易形成拍音。

图 3.21 表示以 A4=440Hz（中央 C 上方 A）为标准参考频率的十二平均律中，各个音符基频 f_0 的大小，同时对中央 C 两侧 8 个倍频程的各个音符进行命名。十二平均律系统主要应用于现代键盘乐器中，但是现在越来越多的表演者和听众对使用不平均律有着浓厚兴趣。不平均律主要应用在古乐器和电子合成器中，电子合成器可以采用各种调律系统。Padgham（1986）对英国大约 100 个管风琴进行了统计，包括 McClure 管风琴在内，这 100 个管风琴都采用了不平均律。

[注 1] 一个音分等于半音的$\frac{1}{100}$。

[注 2] 在现代平均律音阶中，除了倍频程以外任何音程的频率比都不满足整数比关系。

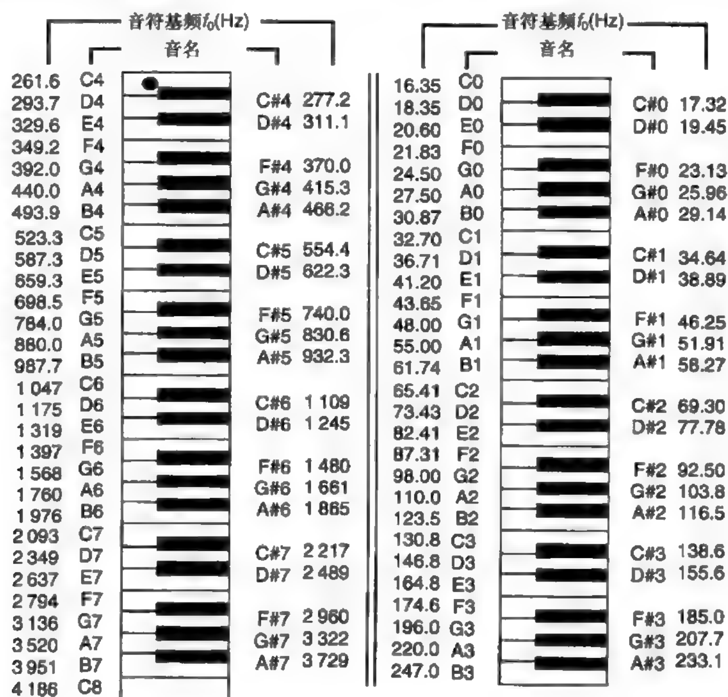


图 3.21　十二平均律中央 C 两侧共 8 个倍频程的各个音符基频 f_0 值（中央 C 用黑点标注，A4=440Hz 为参考频率）

参考文献

Howard, D.M., Hirson, A., Brookes, T. and Tyrrell, A.M. (1995).Spectrography of disputed speech samples by peripheral human hearing modelling. Forensic Linguistics, 2, (1), 28–38.

Moore, B.C.J. (1982). An Introduction to the Psychology of Hearing. London: Academic Press.

Moore, B.C.J. (1986). Frequency Selectivity in Hearing. London: Academic Press.

Padgham, C.A. (1986). The Well-tempered Organ. Oxford: Positif Press.

Parkin, P.H. (1974). Pitch change during reverberant decay. Journal of Sound and Vibration, 32, 530.

Pickles, J.O. (1982). An Introduction to the Physiology of Hearing. London: Academic Press.

Plomp, R. and Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. Journal of the Acoustical Society of America, 38, 548.

Roederer, J.G. (1975). Introduction to the Physics and Psychophysics of Music. New York: Springer-Verlag.

Rossing, T.D. (1989). The Science of Sound. New York: Addison Wesley.

Schouten, J.F. (1940). The perception of pitch. Philips Technical Review, 5, 286.

Wever, E.G. (1949). Theory of Hearing. New York: Wiley.

Zwicker, E., Flottorp, G. and Stevens, S.S. (1957). Critical bandwidth in loudness summation. Journal of the Acoustical Society of America, 29, 548.

第 4 章　乐器的声学模型

4.1　乐器的“黑盒子”模型

为了使各种乐器的声学特性能够很好地得到说明并且真正为读者所理解，本章将介绍一种简单的模型。该模型主要用来解释弦乐器、管乐器、打击乐器以及歌声的声学特性。书中 CD 的第 8 ~ 第 61 条提供了在消声室录制的不同乐器演奏的声音，以供参考。任何声学乐器都由两大部分构成：声源和声音修饰器（译者注：声源是指乐器的振动源，声音修饰器指乐器的其他部分）。

在我们所描述的简单模型中，声源称为“输入”，而声音修饰器称为“系统”，输入经过系统后产生的结果称为“输出”。完整的“输入—系统—输出”模型如图 4.1 所示。



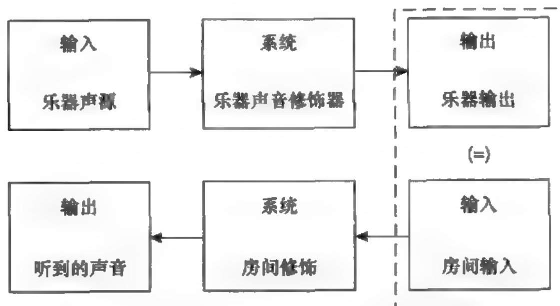
图 4.1　描述乐器声学特性的“输入—系统—输出”模型

这种模型为较好地说明、回顾和理解各种乐器的声学特性提供了一

个有用的框架。请注意这里所说的输出是指真正的乐器输出，而不是人们听到的输出。因为听到的输出是经过了乐器演奏所在房间声学特性修饰了的。下面将阐述怎样将“输入—系统—输出”模型扩展到包含乐器演奏环境的声学特性。

当需要对乐器在某个房间里的演奏效果进行建模时，输出应该是人们所听到的声音，而不是乐器本身产生的输出。房间成为了另一个声音修饰器，因此，在“输入—系统—输出”模型中它应看成是系统的一部分，声学环境模块的输入应该是乐器的输出。因此，乐器在房间中演奏的完整且实用的“输入—系统—输出”模型如图4.2所示，其中乐器的输出就是房间的输入。

图4.2 在房间演奏的乐器“输入—系统—输出”模型



为了使模型具有一定的实用性，必需掌握“输入”和“系统”模块的更为详细的声学特性，这样才能最终确定“输出”。房间的作用将在第6章介绍。本章将介绍弦乐器、管乐器、打击乐器以及歌声的“输入”和“系统”特性。这些详细的声学信息可以通过对其工作原理进行理论分析得到，也可以通过实验测试得到。在实验测试中，应保证录音环境对录制的声音没有影响，或者在有影响的情况下，其影响的大小已知并可以用数学式子反映出来。所谓对声音没有影响的环境是指房间里不存在反射声，其理想状态称为“自由空间”。实际上，自由空间可以在称为消声室的声学实验室里获得。在消声室里，所有到达侧墙、地面和天花板的声音完全被尺寸较大的尖劈状的吸声材料所吸收。然而，消声室并不容易找到，在实际应用中可以用无风天气下广阔的草坪来近似自由空间，这时要求录音或测试在离开地面合理的高度进行。

本章将介绍弦乐器、管乐器和打击乐器的声学特性，其中不同类型

乐器的声源特性以及声音修饰器特性将分别进行介绍。下面的讨论并不是对所有乐器进行分析，而只是选择一两种乐器作为例子，说明如何利用上述的声源和声音修饰器模型来分析乐器的声学特性。对于那些希望对某件特殊乐器的声学特性有更多了解的读者，可以从参考文献中找到有关资料。

歌声的声学特性将在本章最后进行讨论。通常一个没有多少声学知识的音乐技师能够用噪音很好地模拟出电子乐器或声学乐器发出的声音，如果他们具备一些人声的基本常识，那么这件事情做起来会更加容易。作为讨论乐器声学的起点，交响乐中使用的各种乐器以及管风琴、钢琴、歌声的基频范围如图 4.3 所示，图中 9 个八度音程的键盘作为音高参考，其中带黑色方块的琴键代表中央 C。

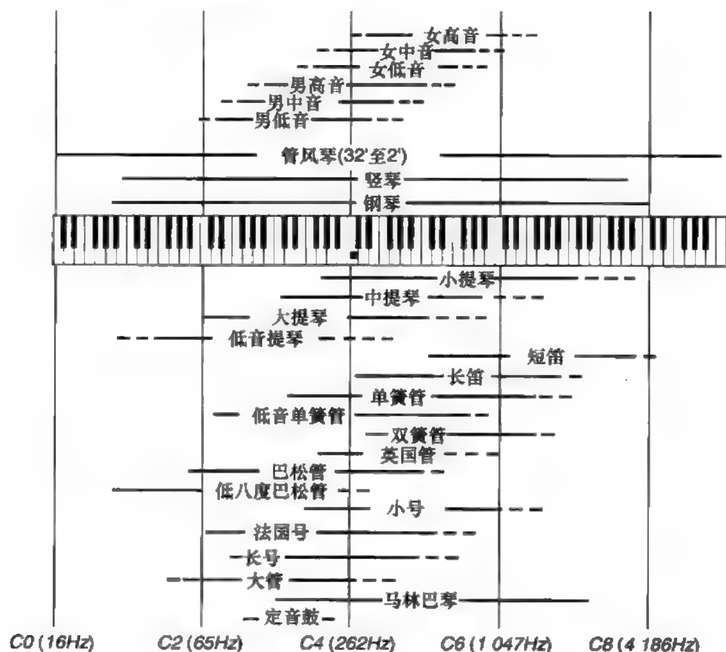


图 4.3　部分声学乐器和歌声的基频范围

4.2　弦乐器

弦乐器家族包括小提琴、中提琴、大提琴、低音提琴和其他较早时

期的弦乐器，以及一些由弦发声的键盘乐器如钢琴、拨弦键琴、击弦键琴（钢琴前身）和古钢琴等。任何一种弦乐器的输出都可以认为是由乐器的输入经过乐器本身声音修饰器后所产生，如图 4.1 所示。更详细的关于弦乐器的介绍可以在参考文献 Hutchins(1975a, b)、Benade(1976)、Rossing(1989)、Hall(1991) 以及 Fletcher 和 Rossing(1999) 中看到。管弦乐队中使用的弦乐器的基频范围如图 4.3 所示。

所有的弦乐器都包含一根或一根以上两端固定并张紧的弦，每根弦的基频 f_0 由弦的线密度（单位长度的质量）、弦长和弦的张力共同决定。在实际乐器中，一般每根弦的线密度是固定不变的，而不同的音高是通过改变弦长或张力获得的。图 4.4 所示为一根一端固定的弦，跨过两个琴马并通过滑轮，另一端悬挂质量可变的重物，弦的张力可通过改变重物的质量来改变，弦长可通过移动右边的琴马来改变。在实际弦乐器中，每根弦的张力通常通过旋转缠绕弦的栓子来调整，从而达到调音的目的，并通过调整其中一个支点的位置即弦长来演奏不同的音符。但是，钢琴等一些弦振动键盘乐器属于例外，它们的每根弦长是固定的，不同的音符由不同的弦产生。


图 4.4　弦长和张力可调的理想化琴弦

振动的弦主要作为乐器的振动源。由于弦的尺寸较小，只有很少的振动能量传递给周围空气质点，因此弦本身的振动几乎不能发出任何声音。任何实用的弦乐器都包含一个共鸣体，弦振动以某种方式传递给共鸣体。由于共鸣体有较大的振动面积，因此可以将较多的振动能量传递给周围的空气。乐器的共鸣体即乐器的声音修饰器，它将本身的振动特性附加到由弦振动产生的声学输入信号上，如图 4.5 所示。

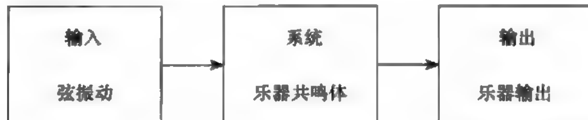


图 4.5　弦乐器的“输入—系统—输出”模型

弦乐器中将能量传递给弦的方式有 3 种：拨弦、弓弦和击弦。同时采用拨弦和弓弦方式的乐器包括提琴家族；仅采用拨弦方式的乐器包括

吉他、诗琴（译者注：主要在 14 ~ 18 世纪的欧洲使用的梨形拨弦乐器）和拨弦键琴；钢琴是一种击弦乐器。

图 4.4 所示的跨过两个琴马张紧的、两端固定的弦的振动是由一系列特有的驻波组合而成的（有关驻波的概念参看第 1 章）。弦在某一时刻的振动形态可以看成由一部分或全部的这些驻波振动模式叠加而成的，或者说，弦振动可以分解为一系列这样的驻波振动模式。两端固定弦的前 10 个振动模式如图 4.6 所示，其中每个振动模式展示的是弦上各点的振幅随两端点之间位置变化的特性。由于弦的两端是固定的，因此每个振动模式在两端的位移都为零，这些点称为“波节”；而有些点的位移振幅最大，这些点称为“波腹”。由图 4.6 可知，第一个模式有两个波节（在两端点处）和一个波腹（在中心点）；第六个模式有 7 个波节和 6 个波腹。一般来说，两端固定弦的第 n 个振动模式有 $n+1$ 个波节和 n 个波腹。每个驻波模式的振动频率与弦长以及弦的横波传播速度有关，见式（1.20）。弦的横波传播速度见式（1.7）。

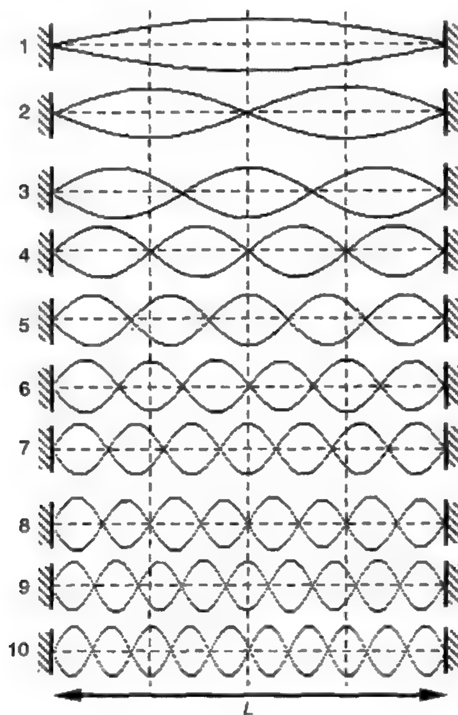


图 4.6　两端固定、长度为 L 弦的前 10 个振动模式

4.2.1　拨弦声源的特性

当拨动一根琴弦时，弦会离开平衡位置产生一个小的位移，然后被释放。由拨弦产生的、提供给乐器共鸣体的声源特性，部分与拨弦的位置有关，因为弦所产生的振动模式多少与拨弦的位置直接相关。例如，当一根弦在中心位置被拨动时，参看图 4.6 中的中心垂直虚线，则在中心点为波节的振动模式（第 2 次、第 4 次、第 6 次、第 8 次、第 10 次等偶次振动模式）就不会被激发出来，而那些在中心点为波腹的振动模式（第 1 次、第 3 次、第 5 次、第 7 次、第 9 次等奇次振动模式）就会被最大程度地激发出来。如果拨弦位置在距离任一端 1/4 弦长处（参看图 4.6 中的其他垂直虚线），则在该点为波节的振动模式（第 4 次、第 8 次等）不会被激发，而其他的振动模式会或多或少地激发出来。总之，当两端固定的弦在距离较近一端 d 处受到拨动时，那些在该点为波节的振动模式就不会被激发出来，它们的次序是：

$$\text{不被激发的模式} = m\left[\frac{L}{d}\right] \tag{4.1}$$

其中，m 为 1，2，3，4…正整数，L 为弦长，d 为拨弦点到较近端的距离。

因此，如果拨弦点在距端点 1/3 处，则不能被激发的模式是第 3 次、第 6 次、第 9 次、第 12 次、第 15 次等。值得注意的是，要使某个振动模式完全不被激发，则拨弦点的距离应该是弦长的某个整数分数，使得该模式的波节点与拨弦点正好重合。

这就是拨弦乐器提供给共鸣体的声输入。两端固定弦的各种模式的振动频率 f_n 与弦长、张力 T 和弦的线密度 μ 有关，将计算横波波速的式（1.7）代入式（1.20）得：

$$f_n = \left[\frac{n}{2L}\right]\sqrt{\frac{T}{\mu}} \tag{4.2a}$$

其中，n 为 1，2，3，4，…，L 为弦长，T 为张力，μ 为弦的线密度。

最小的振动频率由式（4.2a）取 n 等于 1 时得到：

$$f_1 = \left[\frac{1}{2L}\right]\sqrt{\frac{T}{\mu}} \tag{4.2b}$$

该频率就是弦的基频 f_0，也称为第一次谐波频率（见表 3.1）。式（4.2）表明，高次模式的振动频率与 f_0 成谐波关系。

4.2.2　击弦声源的特性

钢琴是一种以击弦作为声源的弦乐器。对于击弦声源而言，前面所讨论的关于缺失某些振动模式的理论同样适用。钢琴的特殊之处在于，其弦的张力要比拨弦键琴或其他交响乐中使用的拨弦乐器的张力大得多，因此显得十分坚硬。钢琴上的琴弦由琴锤进行击打，琴锤击打琴弦后迅速弹回，以免干扰琴弦的自由振动。由于钢琴的琴弦比较坚硬，因此，琴弦被击打后，其振动特性在某些方面与棒振动有相似之处，结果使其各次振动模式的频率略有提升，产生了一定的“不谐和”效应，并且这种效应在高次模式表现得较为明显。

式（4.2）在假设弦为理想的弦即截面直径几乎为零的条件下才成立。将式（4.2b）代入式（4.2a），可得高次模式频率和第一次模式频率的关系为：

$$f_n=nf_1$$

实际上，理想的弦是不存在的，每根弦都具有一定的半径，因此对式（4.2）修正后得到式（4.2c）。说明实际各次振动频率并不完全是基频的整数倍，这就产生了一定程度的不谐和性。

$$f_n=nf_1\left[1+\left(n^2-1\right)\frac{\pi^3 r^4 E}{8TL^2}\right] \tag{4.2c}$$

其中，f_n 为第 n 次振动模式的频率，n 为 1，2，3，4，…，r 为弦半径，E 为杨氏模量，T 为张力，L 为弦长。

由上式看出，不谐和性随振动模式阶次的平方（n^2）、弦半径的 4 次方（r^4）的增大而增大，随弦的张力、长度的平方的增大而减小。为了减小不谐和性，可以使弦尽量细（r 小）、长（L 大），并使之处于高度张紧状态（T 大）。如果仅仅通过使弦变粗（r 大）来达到增大低音弦的质量从而降低频率的目的，那么，由于频率的偏离与弦半径的 4 次方（r^4）成正比，低音弦的不谐和效应就会显著增加。因此，许多弦乐器包括钢琴、吉他和小提琴等，其低音弦往往用细线缠绕来增大其质量，同时保持其

中的弦的半径 r 较小。（关于钢琴的声学特性的详细资料可参看参考文献的 Benade 1976、Askenfelt 1990、Fletcher 和 Rossing 1998。）

钢琴一般采用十二平均律音阶（见第 3 章），两个音程间隔为八度的音符的调音规则是使它们同时发声时产生的拍音最小。当进行间隔八度音程的两个音符的调音时，产生拍音的最大可能性是在较高音符的一次谐波和较低音符的二次谐波之间，因此在调音时往往要使这两个频率尽量接近以避免拍音的产生。由于低音弦的二次谐波与一次谐波之间存在上述的非谐和性，低音弦的基频 f_0 应比高音弦的 1/2 基频值稍低一些。

例 4.1　钢琴的某个音符频率为 400Hz，由于非理想弦使其二次谐波频率为 801Hz。为了避免产生拍音，试问比它高一个八度的音符音调应偏高多少音分?

为了避免产生拍音，高八度音符的基频应为 801Hz，比 400Hz 的二次谐波 800Hz 略高一些。利用附录 2 中的式（A2.4），可将频率比（801/800）转换为音分数：

$$音分数 = 3\,986.313\,7\lg\left[\frac{801}{800}\right] = 2.16\ \ (\text{cents})$$

钢琴琴弦的不谐和性随着琴弦的变短而增加。因此，上述例子中反映的八度音程扩展的现象在高音部比较明显。扩展效应一般以中央 C 作为参考点，在音高与中央 C 偏离越远的音域，这种扩展效应越明显。图 4.7 所示为某一小型钢琴键盘上各音高偏离平均律的情况。由图可知，较高的音和较低的音分别调音到高于或低于（八度音程的频率比为 2 ： 1 的）理想的平均律体系，而且以 C 为参考点时，在 C8 处频率扩展效应为提升约 35 音分，在 C1 处为降低约 35 音分。

钢琴键盘一般从 A0（27.5Hz）到 C8（4 186Hz）一共有 88 个琴键，演奏的音域恰好为略大于 7 个八度音程（见图 4.3）。为了减小琴弦的非谐和性，通常使用尽量细的琴弦，这就使得琴弦传递给琴体的能量很小。因此，在大部分音域，钢琴的一个音使用多根琴弦。用于音乐会演奏的三角钢琴的 88 个音符一共使用了 240 根琴弦，低音部使用的是单根缠绕弦，更高的一些音则使用两根缠绕弦，其余的音则使用 3 根弦，其中较

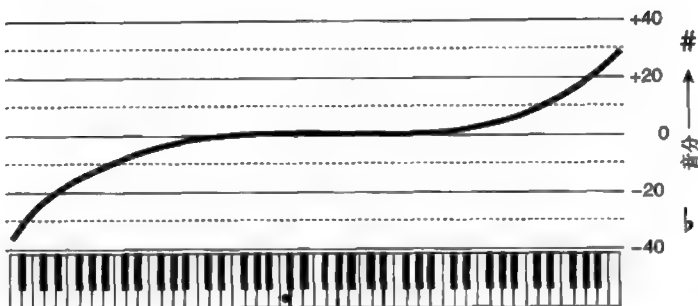


图 4.7　小型钢琴键盘上各音由于不谐和性偏离平均律的音分值（黑点代表中央 C，数据源于 Martin 和 Ward，1961）

低的一部分音仍然使用缠绕弦。多根弦的使用可以控制声音的衰减时间。如果某个音的多根弦（两根或 3 根）能够精确调音而不产生任何拍音（见 2.2 节），则振动能量就会同相地快速传递给音板。多根弦调音的最佳状态是使弦之间存在几个音分的误差，这样会使音色更加丰富，而且声音也不会像多根弦精确调音时衰减得那样快。如果多根弦之间的调音误差达到 12 音分或超过 12 音分，那么这种声音更适合于在酒吧里演奏。

4.2.3　弓弦声源的特性

用弓拉弦产生的声音是周期性的，并且当弓保持向同一方向运动时，声音会持续一段时间。一支弓一般使用多缕毛发，传统上使用马尾毛。弓毛具有方向性，沿一个方向光滑而另一个方向则不然，这一点可以用自己的头发来验证。用一只手抓住头发的末端，另一只手的拇指和食指握住头发的中部，然后上下滑动你的手指，你会感到手指沿一个方向是无法移动的，而沿另一个方向则可以光滑地移动。

琴弓的手握端称为弓根，另一端称为弓尖。弓毛约一半是沿弓根到弓尖方向安装的，另一半则沿相反方向即弓尖到弓根的方向安装。这样不论琴弓往哪个方向移动（上弓或下弓），弓毛都能抓住琴弦。松香常被用于弓毛来增加弓弦之间的抓力。当琴弓沿着琴弦横向移动时，弓毛的抓力使琴弦离开平衡位置，直到琴弦自动脱离琴弓返回并通过平衡位置向相反方向移动，然后琴弓会再次抓住琴弦，使之在平衡位置附近做往返的周期性振动。图 4.8 所示为当琴弓往某一方向移动时琴弦的一个完整周期的运动状态（当琴弓向另一个方向移动时，图形正好相反）。当琴弦被弓毛抓住时，

它以匀速离开平衡位置，当它脱离弓毛时，则以较快速度返回平衡位置直到再次被弓毛抓住。当你用频闪观测仪近距离观察弓弦运动的每个瞬间时，可以看到它是由两个直线区域组成，其交点或分割点沿着图 4.8 所示的一系列快照中的点画线轨迹匀速运动。完成一个周期的时间，即基频对应的周期（T_0）为两个直线区域的交点移动两个弦长距离（$2L$）所需的时间：

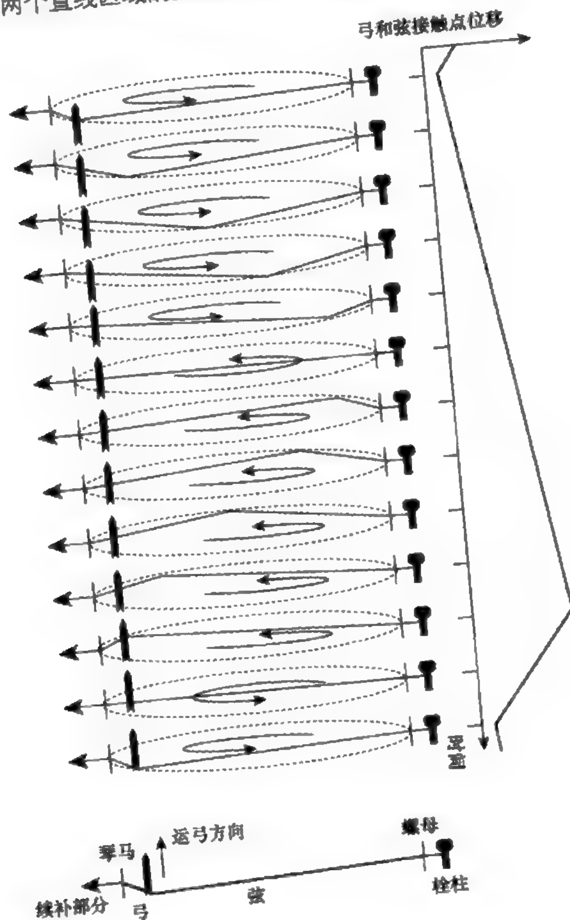


图 4.8　弓弦在一个完整周期的振动状态以及弓和弦接触点位移随时间变化图形（Rossing, 1989）

$$T_0 = \frac{2L}{v}$$

将计算横波传播速度的式（1.7）代入得：

$$T_0 = 2L\sqrt{\frac{\mu}{T}}$$

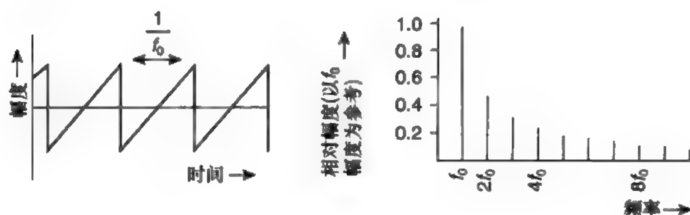
因此，弓弦振动的基频 f_0 为：

$$f_0=\left[\frac{1}{T_0}\right]=\left[\frac{1}{2L}\right]\sqrt{\frac{T}{\mu}} \tag{4.3}$$

将上式与取 $n=1$ 时的式（4.2a）相比，看出这一频率就是弦的一次振动模式频率。由此可见，用弓拉弦是激发弦振动模式的有效方式。

由弓激发的弦振动声源具有弓弦振动的波形，并将振动传递给琴马。图 4.8 中的每一个振动状态对应于图中右边激发点位移波形的相应时间段，由此产生的作用于琴马的力可以认为具有类似的波形。在理想情况下，此波形为锯齿状（见图 4.9）。理想锯齿波的频谱包含所有的谐波成分，其大小随频率的增大以 $1/n$ 的比例减小，其中 n 为谐波的次数。理想锯齿波频谱如图 4.9 所示，各频率成分的大小以 f_0 处大小为参考值。

图 4.9 弓弦产生的理想锯齿波及其频谱



4.2.4 弦乐器的声音修饰器

拨弦或弓弦提供的声源通过琴马传递给乐器的声音修饰器。琴体所有部件的振动特性都对声音的修饰起一定作用。对于提琴家族而言，对声音修饰起最大作用的是面板（在弦的正下方，琴马放置其上，并带有 *f* 音孔）、背板（琴的背部）和琴体内的空腔，其余部分对声音的修饰作用相对要小得多。关于传统提琴制作中所使用的胶水、清漆、木材的选用和一些特殊处理方法对提琴音质的重要性问题，在某些地区还存在激烈的争议。

在提琴家族中，琴体对声音的修饰作用主要表现在低频的两个共振：一是由琴体内空气产生的共振，或称为空腔共振；二是由面板产生的共振，或称为面板共振。Hall（1991）对提琴琴体的共振特性做了如下总结：

（1）实际上在频率低于大约 273Hz 的第一谐振频率（空腔共振频率）时几乎不存在输出；

（2）另一个显著的共振出现在约 473Hz（面板共振）；

（3）在低于 900Hz 的频率范围，频率响应特性很不平坦，特别是在 600 ~ 700Hz 会出现一个较深的谷点；

（4）在 900Hz 以上的频率范围，由于存在较好的振动模式交叠现象，频率响应比较平坦（也存在个别例外情况）；

（5）在更高的频率范围，响应逐渐减小。

除了空腔共振特性主要由空腔的尺寸、形状和 f 音孔的形状等因素决定以外，提琴这类乐器的频率响应与其面板和背板的主要振动模式有关。当制作者在对面板和背板进行仔细雕刻定形时，他们往往会握住琴板的某个位置敲击，通过聆听敲击声了解琴板的振动特性，从而完成琴板的定形制作。这个技术是乐器制作工艺中非常重要的部分，有经验的制作者能够由此了解到，当这些部件装配起来后，乐器具有怎样的共振特性。

乐器最终的声音是乐器声源经过乐器本身的修饰作用后产生的输出。图 4.10（源于 Hall，1991）所示为 G3（f_0=196Hz）弓弦振动的频谱（声源频谱）、典型的小提琴频响特性（系统频谱）以及输出声音的频谱。值得注意的是，由于频率采用对数坐标，因此，输入和输出频谱中的高频成分之间的间隔变小了。输出频谱是由输入频谱乘以琴体的频率响应得到。但是，当幅度大小用对数值即分贝值表示时，乘法运算就变成加法运算。图中所示的输出频谱就是输入的分贝值与频响特性的分贝值相加的结果。

交响乐队提琴家族（小提琴、中提琴、大提琴和大贝司）的成员之间还存在一些基本差别，在声学上主要表现为琴体的尺寸相对于空弦的基频 f_0 变得越来越小（Hutchins，1978）。空腔和面板的共振频率与空弦的振动频率之间的关系如下：小提琴的空腔和面板的共振频率分别与 D4 弦（第 2 弦）和 A4 弦（第 3 弦）的空弦振动频率 f_0 相等；中提琴的空腔和面板的共振频率分别与 G3 和 D4（第 2 弦和第 3 弦）中间的频率、D4 和 A4（第 3 弦和第 4 弦）中间的频率大致相等；大提琴的空腔和面板的共振频率分别与 G2 弦的 f_0、D3 和 A3（第 3 弦和第 4 弦）中间的频率大致相等；大贝司的空腔和面板的共振频率分别与 D2 弦（第 3 弦）和 G2 弦（第 4 弦）

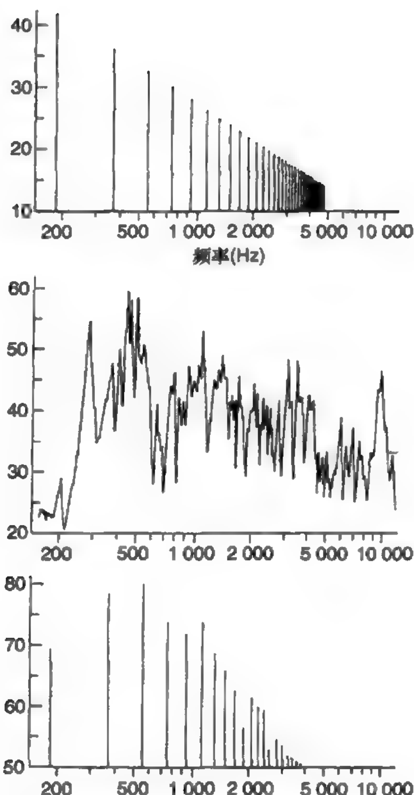


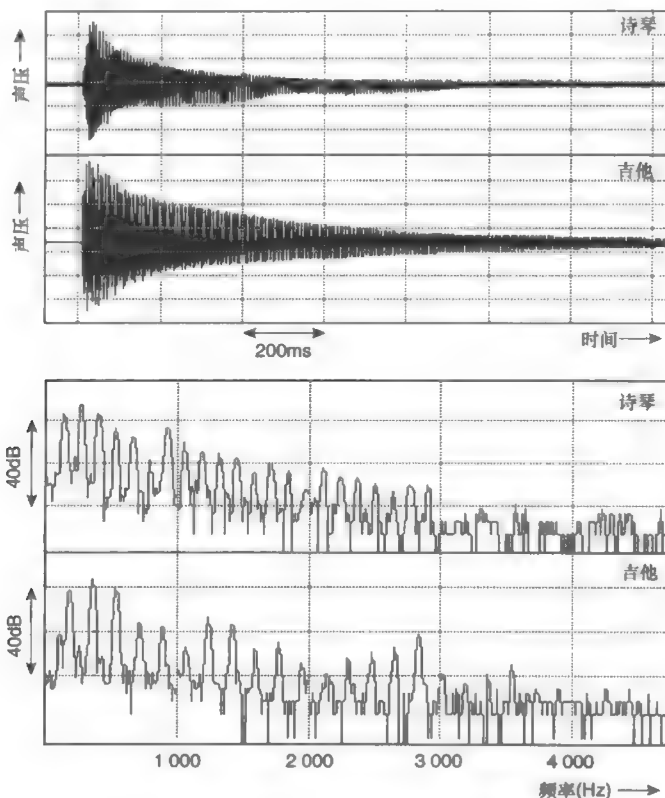
图 4.10 弓弦振动 G3（f_0=196Hz）的频谱、典型小提琴频响特性以及输出声音的频谱（Donald E. Hall 的《音乐声学》中图 11.9）

的 f_0 相等。因此，小提琴的低频共振修饰作用较中提琴或大贝司强。从声学角度上看，这两个起重要作用的共振频率的分布不同是提琴家族每个成员具有各自音色的原因之一。

图 4.11 所示为两个拨弦乐器音符的波形和频谱，即诗琴的 C3 和吉他的 F3。从图中波形上可以看出音符的衰减特性。两个音符的持续时间都只有 1s 多，并且各频率成分之间成整数比例关系（参看式（4.2））。通过观察频谱上振幅减小的谐波成分可以估计拨弦位置。诗琴的频谱说明其拨弦位置大约在 1/6 弦长位置，因为其频谱在第 6 次谐波和第 12 次谐波处有明显的衰减，但是还可以看到其在第 15 次和第 20 次谐波处也存在较明显的下降。但是，必须注意的是，此频谱为乐器输出声音的频谱，因此受到乐器琴体的修饰作用（即空腔共振和面板共振）。频谱不仅与声源有关，而且受声音修饰器的影响。同时，诗琴的第 15

次和第 20 次谐波的振幅比低次谐波小约 40dB，因此，这些频率成分会受到背景噪声较大的掩蔽作用。吉他的频谱说明其在第 6 次、第 12 次和第 18 次谐波有明显的衰减，因此可以估计其拨弦位置也在弦长的 1/6 处。

图 4.11　诗琴的 C3 和吉他的 F3 的波形和频谱



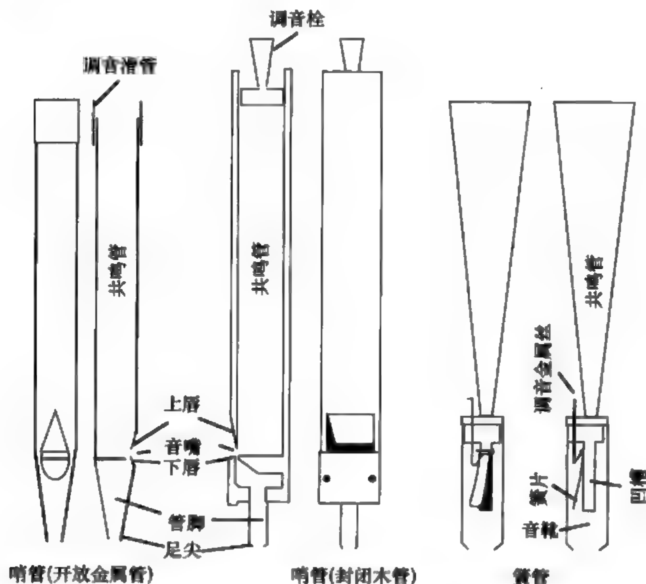
弦乐器的声辐射并不是各方向均匀的。当你仔细聆听或对其进行录音时，会明显感觉到这种差异的存在。不论是拨弦乐器、击弦乐器，还是弓弦乐器，任何一种弦乐器的声输出都包含较宽阔范围的频率成分。一般来说，低频声是向各方向均匀辐射的，从中频到高频，声辐射的方向性变得越来越强。对小提琴而言，无辐射方向性的低频主要是指 500Hz 以下的频率，而高频是指高于 1 000Hz 的频率成分，它们主要是由面板辐射出来。因此，听音位置或录音传声器所在位置对乐声的听音或重放效果有很大影响。

4.3 管乐器

管乐器的分析方法与弦乐器的分析方法类似，管乐器的声源特性与弦乐器有较大区别，而管乐器的声音修饰器特性又与弦的振动模式有共同之处。本节将以管风琴为例介绍管乐器发声的声学机理，其余管乐器的基本声学原理将在本节最后介绍。

管乐器可分为无簧和有簧两类。根据不同的发声机理，管风琴的音管同样可分为“哨管”和“簧管”两类 [注 1]。本节将以管风琴音管为例介绍有簧和无簧管乐器的声学特性。图 4.12 所示为哨管和簧管的基本结构示意图，它们各自都由特别的材料（常用的是木材、锡铅合金）制作而成，具有独特形状和尺寸的共振腔，共振腔形状和尺寸往往由所需产生的声音决定（参考文献 Audsley 1965、Sumner 1975、Norman, H 和 Norman, H.J.1980）。本节首先介绍哨管和簧管的声源特性，然后介绍共振腔产生的声音修饰作用。

图 4.12 哨管（开放金属管和封闭木质管）和簧管结构简图



[注 1] 管乐器和管风琴的发音管可分为有簧和无簧两类。

4.3.1　管风琴哨管的声源特性

Hall(1991) 对哨管的声源特性有详细阐述，这里对其进行介绍。哨管声源的重要构造特征是气流通过的窄缝 (风道) 和对来自窄缝的气流形成阻碍的楔形物。图 4.13 所示为木制哨管的详细构造 (从图 4.12 可看到类似的金属哨管构造)。如图中所示，在哨管的下唇和平板之间有一个窄缝，这个窄缝就是所谓的“风道”。楔形障碍物构成上唇，其位置略微偏离来自窄缝的气流中心线。

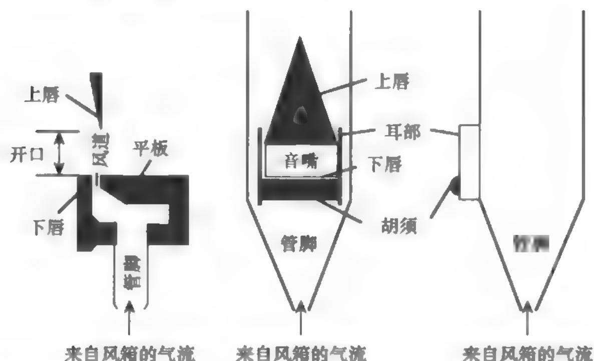


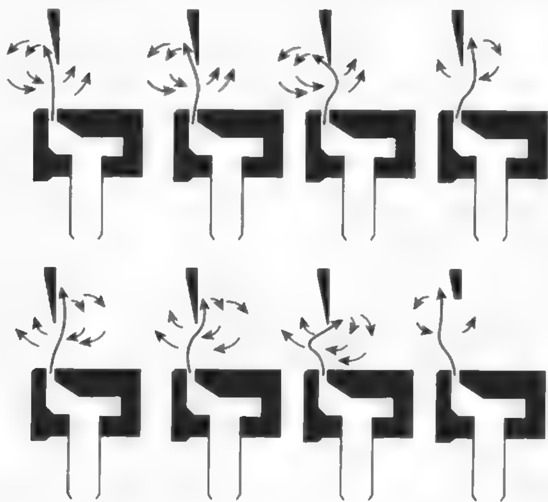
图 4.13　木制哨管声源的基本构造 (左图) 和某些金属哨管的构造特征 (中图、右图)

气流由管风琴底部经过管脚进入音管，在窄缝处产生一个薄的气流层。当不存在上唇时，窄缝产生的声音听起来像噪声，这说明气流处在一种无规扰动状态。当你将上齿放在下唇上发出英文单词“far”中的“ff”音时，就会看到这种现象。由于气流是猛烈而不稳定的，因此，声音听起来是明显的噪声。如果气流的流速被降低，当降低到一定速度时，这种猛烈的扰动就会停止，噪声就不再产生了，这时的气流变得光滑而呈流线形。猛烈而不稳定的气流是产生无调声如英文单词“shoe”中的“sh”和“sea”中的“ss”的物理原因。“sh”和“ss”的波形和频谱如图 3.9 所示。

当楔形障碍物置于来自窄缝的气流中时，就会产生一种清晰的声音，称为“边棱音”。Hall 指出可以用一种简单的方法来展示这种现象，即用嘴正对着一张薄纸片的边缘吹气，即可听到这种声音。虽然研究人员对边棱音的产生机理有不同的认识，但都倾向于认同图 4.14 所示的解释。参看图中的一系列快照形式的图片，边棱音的产生机理可以解释如下：

气流首先流向障碍物的一侧，使得此侧声压增大，将邻近的空气挤出，一部分空气可能形成回流，从音嘴进入音管，同时，回流的速度不断增大，使得来自缝隙的气流改变方向，流向楔形障碍物的另一侧；同样的气流运动过程发生在这一侧，随着声压增大，气流产生回流，从音管内通过音嘴向外流出，并逐渐使主气流改变方向，再次流向楔形障碍物的外侧。这种过程不断重放，从而产生了提供给音管的周期性声源。由于主气流方向不断地左右偏离，这个声音有时可以认为是由“空气簧片”振动产生的。

图 4.14 哨管声源工作原理示意图



在没有共鸣管的情况下，空气簧片产生的脉冲声的基频 f_0 与来自窄缝的气流速度成正比，与音嘴的开口长度成反比：

$$f_0 \propto \frac{v_j}{L_{cut}} \tag{4.4}$$

其中，$\propto$ 表示“正比于”，f_0 为共鸣管不存在时空气簧片振动的基频，v_j 为气流速度，L_{cut} 为音嘴垂直方向开口长度。

换句话说，f_0 可以通过提高气流速度或减小开口长度来提升。当气流速度增大或开口长度减小到一定程度时，f_0 值会增大。这种现象在吹竖笛时会观察到，即当提高吹口处气流速度时，音调会随之提高（如果可能的话，也可以用哨管做同样的实验）。这种吹奏方法通常称为“超吹”。

哨管的声源特性主要取决于音嘴，音嘴决定了各个音管的声音，并能

在各组音管之间建立平稳的音调关系。作为哨管声源，为了适应不同的使用要求，Norman, H 和 Norman, H. J.（1980）对音嘴的调整技巧总结如下：

（1）调整音嘴垂直方向的开口宽度；

（2）切刻平板和下唇；

（3）根据下唇的位置调整平板的高度。

音嘴的调整需要精确地完成，以便在一组音管之间建立稳定的音调关系。对于金属音管，通常使用锋利的、具有短而厚刀刃的刀子完成此项工作。较高的音嘴产生的声音较响、较空，较低的音嘴产生的声音则较柔、较紧。音嘴越高，则要求管脚的气流速度越大。然而，管脚的气流速度越大，则音管声音的瞬态特性越差。

切刻是指在平板和上唇的边缘形成一系列小的切口，目的是减少声源频谱中的高频成分，使音管的声音起振更加平缓。对于重吹的音管需要更多的切刻。没有足够切刻的音管发出的声音具有典型的辅音起振特点，有时称为“合成吹管（chiff）”。管风琴音嘴的调整趋向于采用较少的切刻或无切刻，以便利用合成吹管起振的优势改善乐音的清晰度，这对于复调音乐尤为重要（Hurford, 1994）。

木制音管的平板高度在制作好以后是固定的，而金属音管的平板高度则可以调整。平板可以一定程度地控制哨管气流的方向。当平板太高时，音管发声的反应时间较长，当气流完全不受上唇控制时，音管可能不会发出任何声音；当平板太低时，音管发声的速度太快，声音会变得很难控制。当音管通过少切刻或无切刻方式获得辅音起振特质时，它的发音就具有较快的反应速度。细的音管（管的直径相对于管长小得多）通常具有弦的音色，往往通过增加耳部（见图 4.13）来稳定空气簧片的振动。有些低音管在两个耳部之间安装了木制的滚轴，或称为“胡须”，用来提高音管的反应速度。

4.3.2 管风琴哨管的声音修饰器

管风琴哨管的声音修饰器就是音管本身，即声音的“共鸣管”（见图 4.12）。经过多年的发展，管风琴的音管形成了多种形状，以获得音色的细微差异。然而首先需要考虑的是使共鸣管的截面沿管长保持不

变，即具有均匀的截面积。金属音管通常采用圆形截面，而木制音管通常采用正方形截面（有些制作者采用三角形截面，目的之一是节约原材料）。这些截面形状的采用主要是使各种材料的音管制作起来更加容易些。

有两种基本类型的管风琴哨管，一种是远离哨口的一端是开放的，另一种是远离哨口的一端是封闭的（见图 4.12）。哨口端在声学上可以认为是开放的，因此，开放型哨管的两端都是开放的，而封闭型哨管的一端是开放的，另一端是封闭的。由图 4.14 所示的空气簧片工作原理产生的声脉冲携带能量进入音管。例如，当声压的压缩脉冲（正脉冲），即图 4.14 所示空气簧片在右下图时间段产生的声压信号进入音管时，压缩脉冲将以声速沿管子传播。

当压缩脉冲到达音管的另一端时，它将以第 1 章驻波一节（1.5.7 小节）中介绍的两种方式中的一种方式反射回来，反射方式取决于另一端是开放还是封闭的。如果另一端是封闭的，则在端口会形成一个声压波腹，反射波保持压缩脉冲状态；如果另一端是开放的，则在端口会形成一个声压波节，反射波变为稀疏脉冲状态，以保证管口的压强为大气压强。同样地，当稀疏脉冲到达封闭端时，将保持稀疏状态被反射回来，而到达开放端时，则以压缩脉冲状态反射回来。所有的 4 种情况如图 4.15 所示。

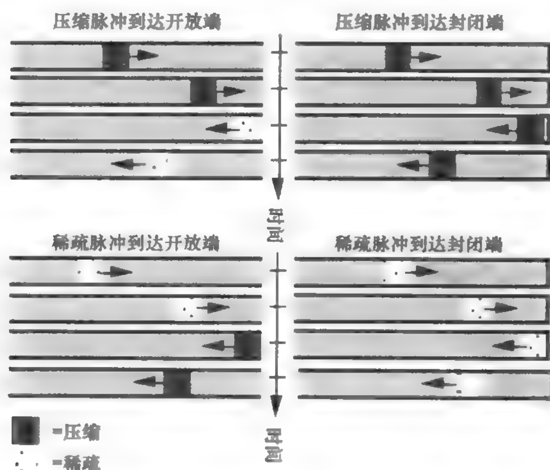


图 4.15 压缩脉冲（上）和稀疏脉冲（下）到达截面均匀管的开放端（左）和封闭端（右）的反射情况（时间坐标轴采用线性刻度）

当考虑到共鸣管对空气簧片声源（见图 4.14）的影响时，声源振动

的基频完全由下面两个因素决定：（a）共鸣管的长度；（b）另一端是开放的还是封闭的。空气簧片振动基频 f_0 对此的依赖性可用声压脉冲到达和离开开放端或封闭端的过程来说明。

图 4.16 所示为一组快照图片，表示空气簧片产生的声压脉冲在长度分别为 L_o 和 L_s 的开放管和封闭管传播的情况以及管长对空气簧片振动基频的影响（空气簧片的振动原理见图 4.14）。图中每个音管的上 1/3 部分显示从左向右传播的压力波，中间 1/3 部分显示从右向左传播的压力波，下 1/3 部分显示两列压力波之和，时间坐标轴采用线性刻度。由于压力波以恒定的速度前进，所以每个等时间间隔传播的距离相等。在考虑声波的反射方式时，音管的哨端可看成是开放端（见图 4.15）。每当压力波到达并从哨端反射时，如果到达哨端的是压缩波并以稀疏波反射回去，则空气簧片由内向外移动；如果到达哨端的是稀疏波，则情况相反，如图 4.16 所示。

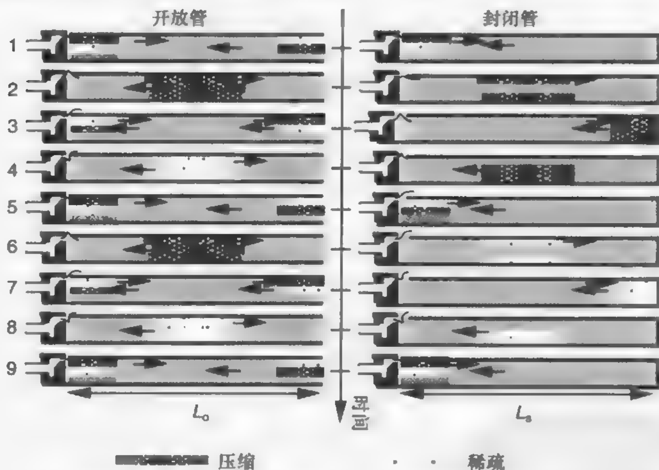


图 4.16　压力波在开放管（左）和封闭管（右）中传播并使空气簧片振动（时间坐标轴采用线性刻度，音管上部显示从左向右的传播，中部显示从右向左的传播，下部显示两列波之和）

对开放管而言，第 1 个图表示一个压缩波从哨端开始向管内传播，同时另一个压缩波正离开音管的开放端（后面很快会解释这个压缩波存在的原因）；第 2 个图表示两列波到达音管中心的瞬间，这时音管中心位置的声压达到最大值，同时两列波保持原有特性继续传播（见第 1 章声波叠加原理）；第 3 个图表示压缩波到达开放端后以稀疏波形式反射回来，而另一个压缩波到达哨端后使空气簧片向外偏移，并产生稀疏波（这看来似乎难以理解，但它是产生第 5 个图的原因）；第 4 个图表示两列稀疏波在中心位置相遇的时刻，中心的声压是两个稀疏声压之和，达到最小

值；第 5 个图表示当稀疏波到达哨端并以压缩波反射回来时，空气簧片由外向内偏移。至此，由于第 1 个图和第 5 个图完全相同，一个周期的振动完成了（为了方便与封闭管的情况进行比较，图中显示了第 2 个完整周期）。

开放管的基频周期就是完成一个完整周期需要的时间（即从一个压缩波离开哨端到下一个压缩波离开哨端的时间）。根据图 4.16 所示，它对应于时间轴上的 4 个时间间隔（第 1 图到第 5 图），即压力波从一端传播到另一端，再传播回来所需的时间，也就是传播两倍管长所需的时间：

$$T_{0(\mathrm{open})} = \left[\frac{2L_{\mathrm{o}}}{c}\right]$$

其中，$T_{0(\mathrm{open})}$ 为开放管的基频周期，L_{o} 为开放管长度，c 为声速。

因此，开放管的基频为：

$$f_{0(\mathrm{open})} = \left[\frac{1}{T_{0(\mathrm{open})}}\right] = \left[\frac{c}{2L_{\mathrm{o}}}\right] \tag{4.5}$$

对封闭管而言，图 4.16 中的第 1 个图也是从压缩波离开哨端开始的，但不存在离开另一端的任何压力波。第 2 个图表示该压缩波传播到音管中心的瞬间；第 3 个图表示压缩波到达封闭端并以压缩波形式被反射的瞬间（见图 4.15），该位置的声压为两个压缩波声压之和，达到最大值；第 4 个图表示压缩波回到中心位置；第 5 个图表示压缩波到达哨端并以稀疏波形式反射回来，同时，空气簧片由内向外偏移；第 6 个图表示稀疏波到达音管中心；第 7 个图表示稀疏波到达封闭端并以稀疏波形式反射回来，并且在封闭端的声压为两列稀疏波之和，达到最小值；第 8 个图表示稀疏波返回到中心位置；第 9 个图与第 1 个图完全相同，说明空气簧片完成了 1 个周期的振动。

显然，当音管的长度相同（并且忽略细管长度实际存在的修正值）时，封闭管在一个周期内传播的距离是开放管的两倍，因此其基频的周期是开放管的两倍，它的基频 f_0 是开放管的 1/2，或者说相差一个八度音程。基频周期是压力波从哨端开始传播两个来回所需的时间，或者说是传播 4 个管长距离所需的时间（见图 4.16）：

$$T_{0(\mathrm{stopped})} = \left[\frac{4L_{\mathrm{s}}}{c}\right]$$

其中，$T_{0(\text{stopped})}$ 为封闭管的基频周期，L_s 为封闭管长度，c 为声速。

因此：

$$f_{0(\text{stopped})}=\left[\frac{1}{T_{0(\text{stopped})}}\right]=\left[\frac{c}{4L_s}\right] \tag{4.6}$$

例 4.2　如果开放式音管和封闭式音管的长度相同，试问它们基频 f_0 的关系是什么?

设 $L_o=L_s=L$，代入式（4.5）和式（4.6）得：

$$f_{0(\text{open})}=\left[\frac{c}{2L}\right]$$

$$f_{0(\text{stopped})}=\left[\frac{c}{4L}\right]$$

因此：

$$f_{0(\text{stopped})}=\frac{1}{2}\left[\frac{c}{2L}\right]=\left[\frac{f_{0(\text{open})}}{2}\right]$$

因此，$f_{0(\text{stopped})}$ 比 $f_{0(\text{open})}$ 低八度音程（频率比为 2 ∶ 1）。

在第 1 章的驻波一节对不同边界细管的自由振动模式进行了介绍。式（1.20）和式（1.21）分别给出了开放管和封闭管（一端开放另一端封闭）各次振动模式频率的计算公式，在上述两个公式中速度都是指声速 c。

由式（1.20）得，开放管的第一次振动频率为（取 $n=1$）：

$$f_{\text{open}(1)}=\left[\frac{c}{2L_o}\right] \tag{4.7}$$

可见，上式与考虑压力波在开放管内传播特性得到的式（4.5）相同。根据式（1.20），其他各次振动频率与 f_0 的关系表示如下：

$$f_{\text{open}(2)}=\left[\frac{2c}{2L_o}\right]=2f_{\text{open}(1)}$$

$$f_{\text{open}(3)}=\left[\frac{3c}{2L_o}\right]=3f_{\text{open}(1)}$$

$$f_{\text{open}(4)}=\left[\frac{2c}{2L_o}\right]=4f_{\text{open}(1)}$$

一般表达式是：

$$f_{\text{open}(n)} = nf_{\text{open}(1)} \tag{4.8}$$

开放管的各次振动模式之间满足谐波关系，并且包含所有的谐波成分。各个谐波之间的音程关系如图 3.3 所示。

封闭管的基频由式（1.21）得（取 $n=1$）：

$$f_{\text{stopped}(1)} = \left[\frac{c}{4L_s}\right] \tag{4.9}$$

上式与考虑压力波在封闭管内传播特性得到的式（4.6）相同。根据式（1.21），其他各次振动频率与 f_0 的关系表示如下：

$$f_{\text{stopped}(2)} = \left[\frac{3c}{4L_s}\right] = 3f_{\text{stopped}(1)}$$

$$f_{\text{stopped}(3)} = \left[\frac{5c}{4L_s}\right] = 5f_{\text{stopped}(1)}$$

$$f_{\text{stopped}(4)} = \left[\frac{7c}{4L_s}\right] = 7f_{\text{stopped}(1)}$$

一般表达式是：

$$f_{\text{stopped}(n)} = (2n-1)f_{\text{stopped}(1)} \tag{4.10}$$

其中，$n=1$，2，3，4，…。

封闭管的各次振动模式之间也满足谐波关系，但是只有奇次谐波存在。各次谐波之间的音程关系如图 3.3 所示。

在开放式哨管和封闭式哨管中，共鸣管是声音修饰器，而声源是空气簧片。空气簧片产生的声音取决于进入音管的压力波波形，而波形又与音管的吹奏方式有很大关系。当进行超吹时，基频提升为第二次模式的频率，即开放式音管的基频提高一个八度音程，为二次谐波频率；封闭管的基频则提高一个八度加一个五度，为三次谐波频率。

空气簧片的振动基频由音管长度决定（见图 5.16）。音管存在自身的振动模式以及相应的振动频率，各次振动的强弱关系由制作音管的材料以及管子截面直径与长度的相对值决定，一般较粗的管子其高次谐波成分较弱。管风琴音管是通过调整共鸣管的长度进行调音的，开放式音管一般是通过管端的滑动音管进行调音，而封闭式音管则是用管端的封闭

塞子（调音栓）进行调音（见图 4.12）。

由于封闭式音管的基频比开放式音管低一个八度，因此，在管风琴空间受限的情况下，常常采用封闭式作为低音音域的音管，并用踏板进行演奏。然而，空间的节省带来的不利因素是声音只有奇次谐波成分。图 4.17 所示为用 Gedackt 8' 和 Principal 8'（译者注：管风琴音栓名称，用来控制使某一组音管发声或不发声）演奏的中央 C 波形和频谱（第 5.4 节将介绍管风琴音栓尺数 8'、4' 等的含义）。Gedackt 音栓控制木质封闭管，频谱显示只有奇次谐波成分，尤其是第 1 次、第 3 次和第 5 次谐波较强；Principal 音栓控制金属质的开放管，其频谱同时呈现奇次谐波和偶次谐波。尽管这两个音栓的音高是相同的，都用 8' 表示，封闭式的 Gedackt 组音管长度只有开放式的 Principal 组音管长度的一半。

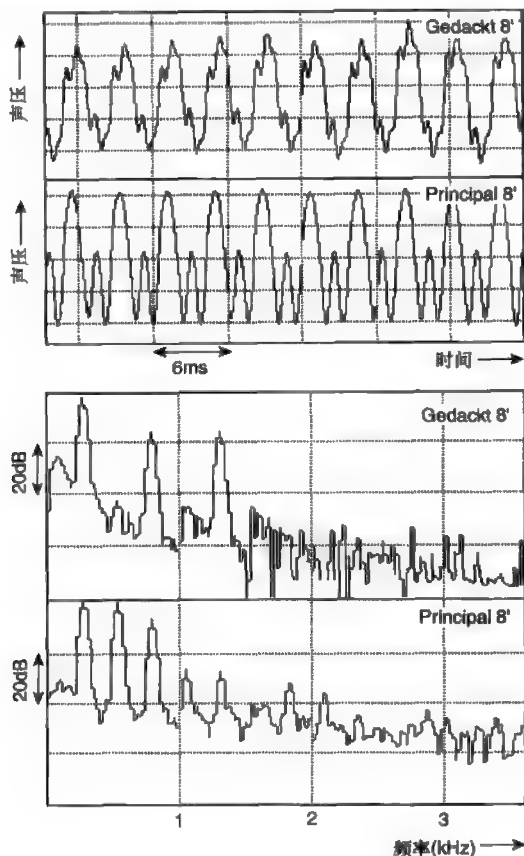


图 4.17 在 Gedackt 8'（封闭式哨管）和 Principal 8'（开放式哨管）演奏的中央 C（C4）波形和频谱

4.3.3 无簧木管乐器

其他采用空气簧片作为声源的乐器包括竖笛和长笛。关于木管哨口乐器的详细资料可参看参考文献 Benade(1976) 和 Fletcher 和 Rossing (1999)。空气簧片的振动是由不断进出哨口的气流控制的（见图 4.16），通常称为"气流控制阀"。因此必然在哨口存在质点位移的波腹和声压波节。哨口端相当于开放端，木管哨口乐器通常在声学上可以看成两端开放的管，如图 4.18 所示。

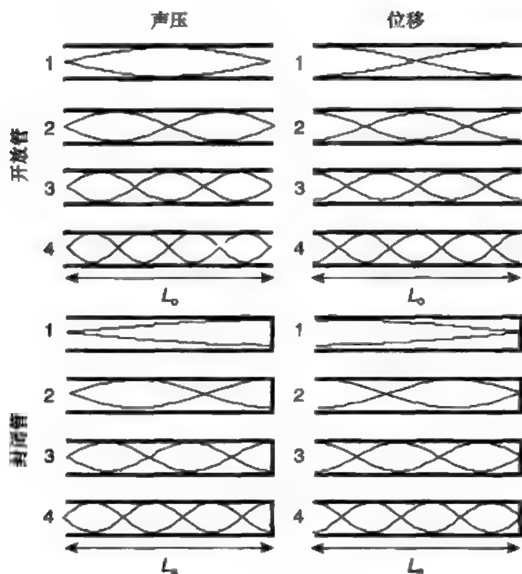


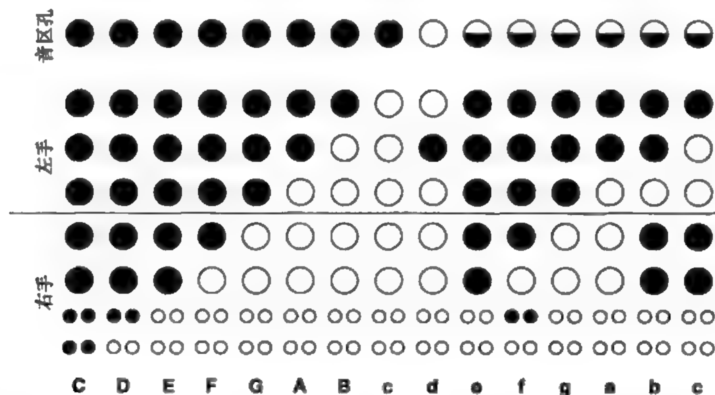
图 4.18 均匀截面开放管和封闭管的前 4 个位移和声压模式（曲线显示声压和位移振幅的最大值和最小值）

演奏者通过改变管长可以用同一件乐器演奏出不同音高的音符。例如，斯旺尼哨（笛）是利用滑动活塞来改变管长的，更常见的是通过手指按住或打开管壁上的音孔来达到改变管长的目的，这些音孔也称为"指孔"。音孔在声学上的作用相当于开放端（对应于声压波节和位移波腹），其近似程度取决于音孔的直径与管子直径的相对大小。当音孔直径相对管的直径较大时（例如长笛的音孔），则没有按住的音孔位置决定了管长且相当于管子的开放端。较小音孔的声学效果是使管长略微向离开哨口的方向延长。这一点对低音乐器的设计存在利用价值。因为低音乐器的

管长较大，通过设计较小的音孔，可以使音孔位置较近并处在演奏者双手能够触及的地方。然而，较小的音孔对各次振动模式频率之间的关系也有影响，后面在关于木管簧片乐器的章节将详细讨论这一点。另一个解决演奏者手不能触及音孔的办法是利用按键，即利用杠杆的机械原理，通过手指按下按键使音孔关闭或打开（取决于音孔在正常状态下被设计成打开或是关闭状态）。

一般来说，当从低到高演奏音符时，一排音孔逐步被打开，使音管的有效长度变短。有时交叉指法会在较小音孔或双音孔乐器例如竖笛中使用，如图 4.19 所示。在此情况下，声压波节的位置会延伸到第一个打开的孔以外，以致后面的音孔状态会影响到声压波节的具体位置。图 4.19 所示为和音（descants）和次中音（tenors）竖笛演奏两个八度的 C 大调音阶的指法。应用各种的指法可以演奏出一个八度中所有的半音音阶，当需要演奏高八度的音阶时，木管乐器如竖笛、长笛等通常采用超吹的办法。由于乐器在声学上是两端开放的管，因此，超吹可以使空气簧片振动频率变成第二次谐波频率，该频率是基频的两倍（见式（4.8）和图 3.3）。在使用超吹法演奏高八度音阶时，可采用相同的指法。

图 4.19　和音和次中音竖笛的 C 大调指法



当需要把音调提高一个八度加五度时，通过超吹可以使基频变为三次谐波频率（与基频的音程差是八度加五度），并运用相同的指法获得更高的音符，第三个八度的开始音就是第 4 次谐波频率，依此类推。在一些乐器如竖笛中，超吹还可以通过打开一个小“音区”孔实现，使某些不需要的振动模式的声压波腹正好处在这个位置，从而达到抑制这些振动模式的目的。图 4.19 所示的音区孔一般位于竖笛的背面，由左手的大

拇指控制，可使其处在完全封闭、半封闭（用大拇指指甲顶部按住音区孔）和完全打开等状态。为了抑制第一次模式，这个孔的位置应处于该振动模式声压最大值处。当所有的音孔都关闭时，这个音区孔应处在整个共鸣管的中心位置。这一点正好是一次振动模式的声压最大值点，因此一次模式被抑制，而这一点又是二次模式的声压最小值点，因此不会影响二次模式的产生（见图 4.18）。音区孔还可以置于其他位置以获得其他模式的超吹。在实际应用中，音区孔位置的选择往往还要兼顾其他方面，使之能够支持该音区的所有音符可以通过打开音孔发声。

长笛的演奏音域在 B3 和 D7 之间，而短笛的音域在 B4 和 D8 之间，比长笛高一个八度（见图 4.3）。长笛和短笛演奏者可通过调整嘴唇与吹口的相对位置来控制超吹模式的稳定性，如图 4.20 所示。空气簧片的工作原理与图 4.13 和图 4.14 所示的管风琴哨管类似，其发声原理也可参看管风琴的哨管部分。长笛演奏者能够沿图中双箭头方向转动长笛来调整上唇与吹口之间的距离，如果引用管风琴的命名法，这个距离称为“开口”（见图 4.13）。此外，通过调整嘴唇之间的形状、宽度和高度可以改变气流的速度以及气流的其他详细物理特征（见式（4.4））。因此，长笛演奏者能够直接对超吹模式的稳定性进行控制。

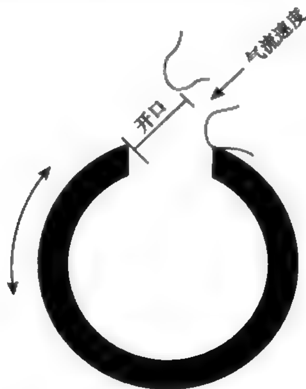


图 4.20 长笛演奏者调整上唇与吹口间距示意图

4.3.4 管风琴簧管的声源特性

管风琴簧管的基本结构如图 4.12 所示。簧管的声源来自簧片的振动，通常簧片比凹槽的开口略大，正对凹槽的边缘。有时也有采用所谓的“自

由簧片”式，即簧片比凹槽开口略小，当簧片向里或向外移动时，不会触及凹槽的边缘。当簧片处在静止状态时，由于簧片不完全是平面，具有一定的弧度，使其与凹槽之间形成微小的缝隙，如图 4.12 所示。簧片的长度可由调音金属丝或调音线圈进行调整。通过轻推使线圈上下移动，从而使振动簧片变长或变短，相应地降低或提高簧片的振动基频 f_0。

当与簧管对应的音栓和键盘上的按键按下时，簧片开始振动发声。参看图 4.12，首先空气进入音靴，并通过簧片和凹槽之间的空隙进入共鸣管。由于簧片与凹槽之间的间隙很小，为了使气流能够进入共鸣管，音靴里的声压必然比凹槽内的声压大，使簧片倾向于关闭空隙，从而使簧片与凹槽的间隙变得更窄；当间隙变窄时，气流的速度变快，同时内外的压强差也变大；增大的压强差向簧片施加一个略微变大的关闭间隙的力；这个过程不断反复，最终使簧片加速向里移动，并快速完全地关闭间隙。

由于簧片具有弹性，当间隙闭合时，气流的速度减小为零，这时簧片的弹性恢复力使簧片向平衡位置方向移动，间隙重新打开。由于惯性，簧片到达平衡位置后继续向外移动，然后停止，再向凹槽方向移动，这就像用手指使簧片离开平衡位置，然后松手，簧片会在平衡位置附近产生振动一样。随后，通过凹槽的气流得到恢复。这个运动过程形成一个周期并不断重复。

当没有共鸣管时，簧片会以其自然（固有）频率振动。自然频率是指簧片被拨动后的振动频率。如果簧片被拨动后振动持续的时间较长，则很可能其包含的振动频率范围很窄；如果簧片振动持续的时间较短，则其包含的振动频率范围较宽，如图 4.21 所示。这种振动持续时间的差异与制作簧片所使用的材料以及簧片的固定方式有关。振动包含的频率范围较窄的簧片往往是由黄铜制成，并且采用刚性固定，称为“硬簧”；而振动包含的频率范围较宽的簧片往往是由芦苇或塑料制成，采用柔韧的固定方式，称为“软簧”。参看图 4.21，自然周期 T_N 与自然频率 F_N 之间的关系是：

$$F_N = \frac{1}{T_N} \tag{4.11}$$

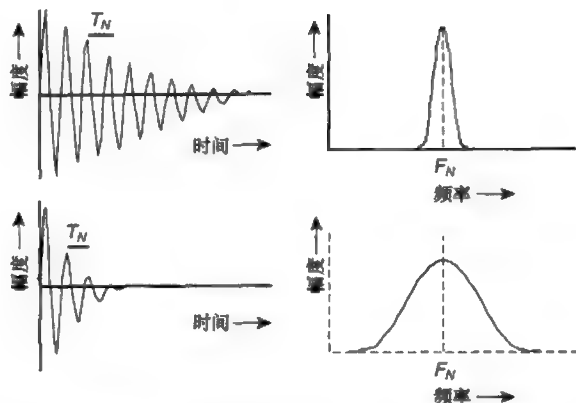


图 4.21 硬簧（上）和软簧（下）拨动后的时间（左）和频率（右）响应（T_N、F_N 分别为自然周期和自然频率）

当簧片向里移动并快速完全地关闭气流通过的间隙时，引起声压的变化，形成提供给共鸣管的声源。气流的快速中断使间隙内的声压急剧变小（因为急速流过间隙的气流突然中断），而波形的快速变化意味着其包含较丰富的高次谐波成分。声源的详细频谱特性是由簧片、凹槽和风箱的压力等因素共同决定的。"自由簧片"振动时不触及凹槽边缘。例如，在口琴、簧风琴中使用的是"自由簧片"，因此不会使气流完全停止，其包含的高次谐波成分就不会占很大比例。

4.3.5 管风琴簧管的声音修饰器

所有的簧管都有共鸣管。共鸣管的作用在介绍哨管时已经用图 4.16 说明了。簧管和哨管的工作原理基本相同，主要区别在于共鸣管的簧片端在声学上应视为封闭端（这一点与哨管的哨端为开放端正好相反）。这是因为在簧片振动时，簧片或者将簧片端完全封闭，或者只留出一个比管径小得多的间隙，因此簧片端应视为封闭端。

管风琴簧管通常使用具有较窄振动频率范围的硬簧（参看图 4.21）。与空气簧片不同，硬簧的振动频率一般不受共鸣管的控制。共鸣管对声源的修饰作用与它所支持的振动模式有关，见图 4.18，但要注意到共鸣管的簧端变为封闭端。由于簧片本身能够适应管子的振动基频 f_0，并不需要共鸣管来特别加强其基频，因此共鸣管往往只加强和支持波长为其长度的整数分之一的高次谐波成分。图 4.22 所示为在 hautbois 8' 或 oboe 8' 和

trompette 8' 或 trumpet 8' 演奏的中央 C 的波形和频谱。两个频谱都在大约第 6 次和第 7 次谐波处呈现峰值；对于 trompette 而言，这个峰较宽，并且一直延伸到大约第 10 次谐波，奇次谐波成分明显地较偶次谐波成分强，这种频谱特征很可能与共鸣管的形状有关；而 hautbois 的频谱比 trompette 呈现更多的谷点，这些都是决定不同乐器具有不同音色的频谱特征。

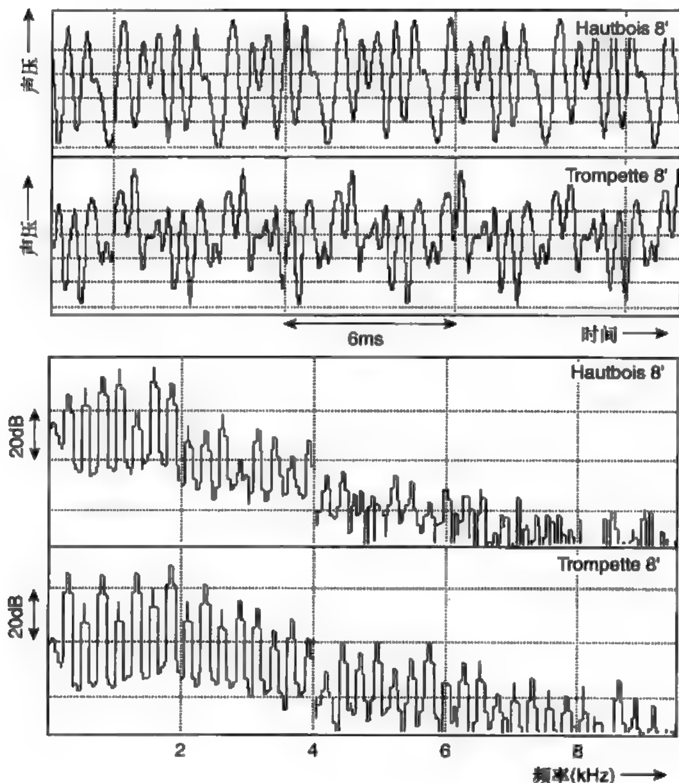


图 4.22　在 hautbois 8' 和 trompette 8' 演奏的中央 C 的波形和频谱

4.3.6　有簧木管乐器

有簧木管乐器分为单簧片和双簧片两种，簧片用于控制从演奏者肺部到共鸣管的气流。簧片的关闭或打开是由其两边的相对压强决定的，因此通常也被描述成"压力控制阀"。音管的簧端应看成是封闭端（声压波腹和位移波节，见图 4.18）。虽然为了使气流不中断，簧片时开时合，但是簧片的开口比管的截面积小得多。因此认为簧端是封闭端是合理的

假设，这一点正好与前面介绍的无簧木管乐器如长笛、竖笛等相反。因为空气簧片是气流控制阀，在哨端产生声压波节和位移波腹，因此哨端应看成是音管的开放端（见图 4.18）。

软簧的振动频带较宽（见图 4.21），常用于木管乐器。单簧管和萨克斯管是单簧片乐器，其簧片相对音嘴打开或闭合，就像管风琴音管的簧片相对凹槽打开或闭合一样；而双簧管和巴松是双簧片乐器。但它们的声源振动机理基本上是相同的。

有簧木管乐器的共鸣管就是声源的声音修饰器，其振动模式对乐器的性能起着决定性作用。木管乐器借助于指孔可以演奏从第一振动模式到第二振动模式之间的任何音符包括半音。当第二次振动模式被激发后，同样的指法仍然适用，这种操作模式使得木管乐器能够演奏出覆盖整个音域的所有半音（见图 4.3）。当音孔打开进入第二振动模式时，保持共鸣管各次振动模式之间的频率比不变是基本要求，否则，进入高次模式后将对调音产生不利影响。Benade（1976）对上述现象进行了总结，并指出其在实际应用的限制：

"对于所有木管乐器来说，当打开音孔进入高次模式时，能保持各次振动模式之间的频率比不变是基本要求，这就对管子的形状有所限制，才能使其适合于作为音乐演奏使用。"

从这个意义上说，适合音乐演奏的管子基本上有两种，一种是圆柱形，常用于如单簧管中；另一种是圆锥形，常用于如双簧管、英国管、萨克斯管和巴松管中。单簧管的圆柱形共鸣管的作用相当于一端（簧端）封闭而另一端开放的管（如前所述），它只支持奇次谐波成分（参看图 4.18），它的基频比同样长度的两端开放的管如长笛的基频低一个八度（参看图 4.3 和例 4.2）。单簧管的第一个超吹模式是第三次振动模式，与第一次振动的音程差是一个八度加一个五度（参看图 3.3）。因此，单簧管和长笛、竖笛等不同，它需要足够多的音孔来完成从第一次振动模式到第三次振动模式之间至少 19 个半音的演奏。

簧端封闭而另一端开放的圆锥形共鸣管却支持所有的谐波成分。Taylor（1976）对其声学作用做了以下说明：

"假设通过某种方法使一个压缩波从较窄一端出发，声波到达开放端后以稀疏波反射回来，这个过程和两端开放的管没有区别。所不同的是，

在稀疏波回到始端的过程中，由于管径变得很小，波速迅速减小，因此实际上几乎没有反射发生。……结果是我们只需考虑一个来回的传播过程，而无需考虑其窄端是开放的还是关闭的。……就其振动模式而言，上述圆锥形管的工作原理与两端开放管更为接近。"

因此，圆锥形共鸣管支持所有的谐波。具有圆锥形共鸣管的乐器有双簧管、英国管、巴松以及萨克斯管家族，它们的超吹模式是二次振动模式，即比基频高一个八度。为了使音调从基频逐步提升到二次谐波频率，需要有足够的音孔来完成其中至少 12 个半音的演奏。

Benade(1976) 把各种形状共鸣管上的一系列打开的音孔统称为"音孔格栅 (tone-hole lattice)"。当存在音孔格栅时，由于音管的有效声学终端位置与频率有关，因此有效音管长度随不同的振动模式而变化。以打开的音孔作为终端的音管长度在低频驻波时较高频时要短一些，因此高频振动模式的频率会持续地略微降低 (波长增大使频率降低)。当频率提高到一定值时，声波不再因为音孔格栅的存在而产生反射，Benade(1976) 称之为"音孔格栅截止频率 (open-hole lattice cut-off frequency)" (质量较好的巴松管此频率大约为 350 ~ 500Hz，较好的单簧管大约为 1 500Hz，较好的双簧管大约在 1 100Hz ~ 1 500Hz)。Benade 注意到这是不同的木管乐器具有不同音色的主要原因，而且与演奏者感觉到的声音的明亮度或暗淡感有很大的相关性。同时还要认识到，关闭的音孔对音管的声学特性也有影响，使音孔所在位置的管径略微变大。下面将讨论其带来的音管声学特性的变化。

为了补偿打开或关闭音孔给各振动模式频率带来的微小变化，可以对音管形状做一些改变。包括使出口成喇叭状、使某些部分逐渐变窄等，或者通过使管子局部放大或缩窄来对声音频率进行小的调整。对于木制音管可以通过挤出或压进蜂蜡的办法来增大或缩窄音管局部的直径 (Nederveen 1969)。放大或缩窄音管局部对各个振动模式频率产生的影响取决于声压驻波的波节和波腹 (或位移波腹和位移波节) 的分布情况。缩窄音管处相对于声压波腹 (位移波节) 位置对频率的影响如下：

(1) 在声压波节 (位移波腹) 处缩窄音管则降低该模式的振动频率；

(2) 在声压波腹 (位移波节) 处缩窄音管则提升该模式的振动频率。

当在声压波节 (位移波腹) 处缩窄音管时，由于缩窄处两端的声压差保持不变，所以气流通过缩窄部分的速度会降低。Benade(1976) 认为

这相当于增大了局部的空气密度。根据第 1 章的介绍，媒质密度的增大会导致声波传播速度的降低（见式（1.1）），由此引起振动模式频率的降低（见式（4.7）和式（4.9））。另一方面，当在声压波腹（位移波节）处缩窄音管时，会使局部的声压增大，从而对叠加形成驻波声波的局部质点振动产生较大的反作用力，这相当于提高该处空气媒质的弹性系数。从第 1 章的介绍可知，提高弹性系数相当于提高杨氏模量，因此使声速提高（参看式（1.5）），从而使振动模式的频率提高（见式（4.7）和式（4.9））。同理，当音管局部放大时，得到的结果正好与局部缩窄的情况相反。

当了解了音管各驻波模式的声压和位移的波节、波腹的分布规律后，就可以对局部放大或缩窄音管后对振动模式频率的影响进行预测。图 4.23 所示为一端封闭的圆柱形音管在沿管长的任何位置局部放大或缩窄后可能使前 3 个振动模式频率产生的变化（根据图 4.18 和图 4.23 可以方便地作出两端开放的圆柱形管的类似图形，感兴趣的读者可以自己尝试着去完成）。

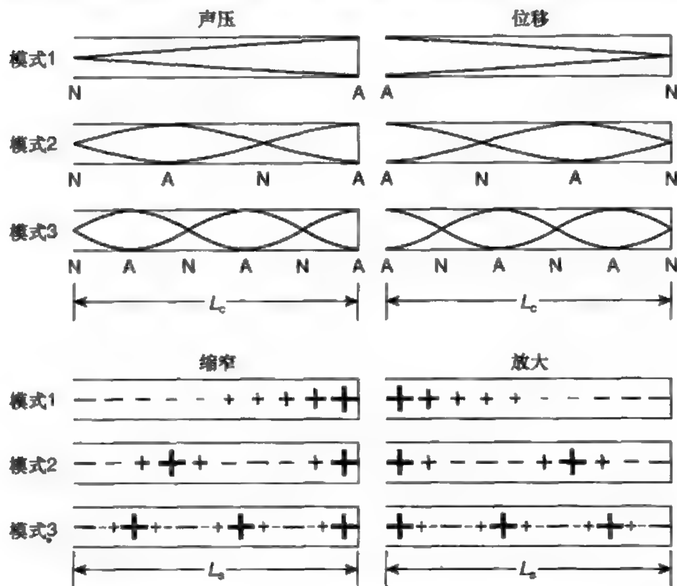


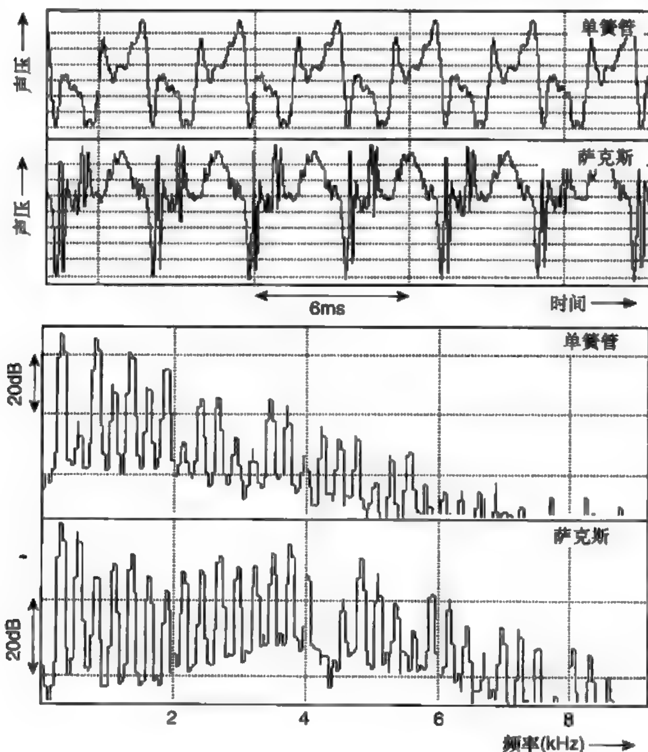
图 4.23 一端封闭的圆柱形音管的局部缩窄或放大对前 3 个振动模式频率产生的影响（“+”表示提高模式频率，“−”表示降低模式频率，“+”、“−”的大小表示频率变化的大小。图中的前 3 个振动模式的声压和位移波节 N、波腹 A 的分布图供参考）

图 4.23 的上部（来自图 4.18）表示一端封闭的圆柱形音管前 3 个驻波模式的声压和位移波节和波腹的位置。图的下部用“+”、“−”表示当沿着整个管长在相应位置缩窄或放大音管时模式频率的增大或减小，符号大小表示其频率变化的大小，而频率变化的大小则取决于缩窄音管的位置与声压或位

移的波节或波腹的间距。间距越小，引起的频率变化越大。例如，在圆柱形管的封闭端缩窄会使所有模式的振动频率提升，因为那是所有模式的声压波腹位置，而在此位置放大音管则会降低所有模式的振动频率；如果在距封闭管的封闭端 1/3 管长处缩窄音管，则第一和第三振动模式的频率略有提升，而第二振动模式频率的降低会达到最大值。熟练的木管乐器制作者通过局部缩窄或放大音管，可以补偿音孔的存在引起的模式频率的偏移，使在整个演奏音域各个模式频率之间保持非常接近于整数比的关系。

图 4.24 所示为单簧管和高音萨克斯管的中央 C 的波形和频谱。因为萨克斯管的共鸣管呈圆锥形，因此其频谱包含所有的谐波成分；而单簧管的共鸣管为一端封闭的圆柱形管（参看图 4.18），因此其频谱呈现较明显的奇次谐波成分，但也可以看出其中包含部分偶次谐波成分。虽然圆柱形共鸣管本身不支持偶次谐波，但由于声源的频谱包含所有谐波成分（萨克斯管和单簧管都是单簧片乐器），因此，单簧管会辐射少量的偶次谐波。

图 4.24　单簧管和高音萨克斯管的中央 C（C4）的波形和频谱



Sundberg（1989）对单簧管的这种声辐射现象作了以下总结：

“这说明共鸣管本身并不支持偶次振动模式。……通常大多数人会错误地认为其输出根本不包含偶次谐波。事实上二次谐波可能会比基波低40dB，因此听觉上很难判断它的存在。在频谱的高频区域，相邻的奇次谐波和偶次谐波之间的大小差异变小。更进一步来说，……这种奇次谐波和偶次谐波之间的差异仅在乐器的低音区存在。”

在图4.24中，单簧管频谱的二次谐波比基波低约40dB。但随着谐波频率的增大，奇偶谐波之间的大小差异变小，这一点与Sundberg的阐述相符合。

4.3.7 铜管乐器

铜管乐器有一个有趣的发展历史。它起源于大自然中管状结构的物体如动物的角、贝壳和植物的茎等，后来演变为木质或金属质乐器，最后发展成为今天管弦乐队成员之一的铜管乐器家族（Fletcher和Rossing 1999、Campbell和Greated 1998）。所有铜管乐器的声源都是吹口上演奏者的嘴唇，它们构成双软簧，演奏者可通过改变嘴唇的张力和形状来调整双软簧的物理性质。就像在有簧木管乐器声源一节中介绍的那样，嘴唇相当于一个压力控制阀。因此，铜管乐器的吹口端在声学上也应等效为封闭端（声压波腹和位移波节，见图4.18）。

当双唇轻微张开，让气流在中间流过时，嘴唇的双簧片作用可作如下解释：当气流速度较小时，什么也听不到；当气流速度增大时，由于空气的扰动，可以听到一些噪声；当气流速度进一步增大时，双唇就开始像双簧片一样振动。由于双唇本身具有一定的弹性以及所谓“伯努利（Bernoulli）效应”（后面会解释）的存在，这种振动会持续下去。

当气流通过窄缝例如双唇间隙时，其速度会增大。伯努利效应是以能量守恒定律为基础的，即在沿着管子的所有位置，动能与势能之和应保持不变。图4.25所示为某一截面伸缩自如的管所体现的伯努利效应（用来说明唇簧的工作原理）。气流方向用带箭头的线段表示，线段之间的间隔反映了流速的大小。当气流通过较窄位置时，由于流速增大，使得动

能增加。为了满足总能量保持不变的伯努利定律，该位置的势能必然减小，这意味着此处对管壁施加的作用力会相应减小。

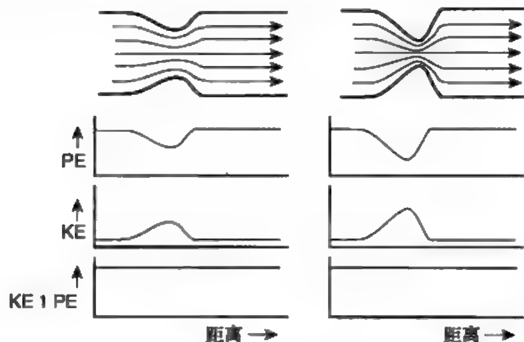


图 4.25　伯努利效应示意图（在局部缩窄的管中，势能 + 动能 = 常数。气流方向用带箭头的线段表示，线段之间的间隔反映了流速的大小。PE 为势能，KE 为动能）

如果缩窄处的管壁具有弹性，并且由伯努利效应产生的力足以使一定质量的物体（例如铜管乐器演奏者的双唇）离开平衡位置，那么管壁会进一步向里收缩一些（比较图 4.25 的左图和右图）；由于缩窄处更窄，使得气流速度进一步增大，动能增大，势能进一步减小来进行补偿，因此管壁加速向里收缩，直到缩窄处完全闭合，切断气流。这时，管内的气压具有使缩窄处重新打开的倾向，管壁也受到恢复力的作用要回到平衡位置，就像两个离开平衡位置的钟摆。管壁回到并通过平衡位置、停止、重新相对运动，伯努利效应加速它们的相对运动，并使它们重新闭合。只要气流不终止，管壁的振动会一直保持下去。如果缩窄处的两个管壁具有相同的质量和张力，例如双唇，那么振动就是周期性的。

唇簧振动由共鸣管支持。铜管乐器的共鸣管通常由一根长管连接到吹嘴构成。通常有一些机械装置使演奏者能够改变管长，最初是用手更换不同长度的音管或弯曲管（例如在号家族中），现在则通过使用伸缩管（例如在长号中），或者使用阀门来增加管长。用伸缩管或阀门调整共鸣管部分的管长时必须采用圆柱形，以便伸缩管或阀门能正常工作。

所有的铜管乐器都由 4 个部分组成（见图 4.26）：吹嘴、截面逐渐变窄的吹嘴管、圆柱形（如小号、法国号和长号）或圆锥形（如短号、粗管短号、中音号和大号）并安装有伸缩装置或阀门的主体音管和喇叭口

（Benade 1976、Hall 1991）。如果铜管乐器的主体音管呈圆锥形，则其支持所有的振动模式（参考前面有簧木管乐器部分）；如果主体音管呈圆柱形，则由于唇簧的压力控制阀作用，应视其为一端封闭的管，因此它仅支持奇次振动模式（见图 4.18）。然而，由于增加了吹嘴和喇叭口，铜管乐器家族几乎都能够支持所有的振动模式，这些振动模式的频率成谐波关系。



图 4.26　铜管乐器的基本结构

喇叭口随着频率的不同，对来自管内的声波反射产生不同的影响。Benade（1976）对此进行了详细讨论，在此对其结论进行总结。低频成分到达喇叭口的较窄部位时被反射回音管里，而高频成分则到达较宽部位时才被反射；当频率高于由喇叭口直径决定的截止频率（小号约为 1 500Hz）时，声波基本上不被反射。因此在乐器出口处增加一个喇叭口具有使有效管长随频率提高而增大的作用，从而使圆柱形封闭管（只有奇次谐波 $1f$、$3f$、$5f$、$7f$ 等）各振动模式频率之间的关系改变，即各频率之间的间隔变小，这种作用对前几个振动模式较为明显。

在主管的另一端加上吹嘴也会对一些振动模式的频率产生影响。吹嘴一般由一个杯形空腔和一个称为吹嘴管的圆锥形短管组成，气流通过杯形空腔的小孔进入圆锥形短管。吹嘴存在一个共振频率，对小号来说大约是 850Hz。由于用手掌击打吹嘴时会听到这个共鸣声，所以这个共振频率称为爆音频率（popping frequency）（Benade 1976）。吹嘴的加入会使有效管长增大，这种效应随着频率的增大而增强。Benade 注意到这种效应随着频率达到乐器的最高音区而稳定地加强，并且对于具有较低爆音频率的吹嘴而言，当频率逐渐提高时会呈现较大的总体长度变化（Benade 1976，第 416 页），因此，吹嘴引起的频率降低效应对高频振动模式的影响比对低频振动模式的影响大。

对于整个铜管乐器来说，通过选择合适的喇叭口、吹嘴和吹嘴管的形状，可以使其振动模式从一端封闭的圆柱形管只支持的奇次谐波系列变为非常接近于包含所有谐波成分的一系列振动模式。在实际应用中，形成的一系列谐波频率与精确的基频整数倍频率只有百分之几的误差，但是第一次振动模式除外。因为它的频率远低于高次谐波的基频，因此它与其他频率之间不是谐波关系。增加喇叭口、吹嘴和吹嘴管对最低 6 个模式频率的影响大致如图 4.27 所示。其中，一端封闭的圆柱形管的最低 6 个奇次模式频率用 f 的整数倍表示，最终形成的铜管乐器的最低 6 个模式频率用另一个频率 F 的整数倍表示。

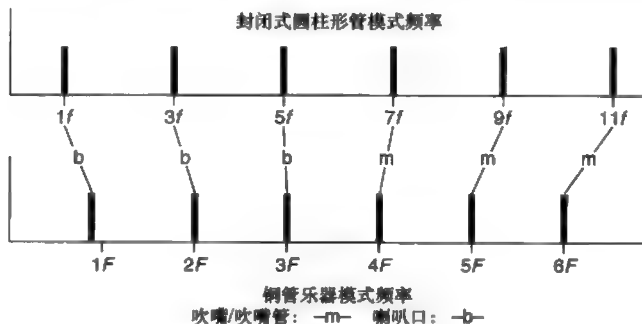


图 4.27　增加吹嘴（和吹嘴管）、喇叭口对封闭式圆柱形管最低 6 个模式频率的修正

因此，第二振动模式是铜管乐器进行音乐演奏可用的最低振动模式（注意这里所说的最低振动模式并不对应于 1F）。当通过超吹使频率从第二振动模式进入第三振动模式时，相应的音调变化是全五度或 7 个半音。铜管乐器（长号除外）采用增加 3 个阀门的方法，可以使其中的 6 个半音得到演奏。当按下每个阀门时，主管长度会有一定数量的增加，使声音能够从前一个振动模式过渡到下一个振动模式。假设全五度音程用 C4 到 G4 表示，则中间缺失的 6 个半音分别是 C#4、D4、D#4、E4、F4 和 F#4。图 4.28 为演奏这 6 个半音的操作示意图。按下中间（或第二个）阀门则降低 1 个半音，按下第一个阀门则降低两个半音，按下第三个阀门则降低 3 个半音。因此，从理论上说，通过这 3 个阀门的不同组合可以完成所需的 6 个半音的演奏。从图中看出，下降 3 个半音可以通过两种指法（仅按下第三个阀门或者同时按下第一个和第二个阀门）获得，这似乎难以理解。下面就来解释有关阀门使用时要遇到的一些重要问题。

半音	−1	−2	−3	−3	−4	−5	−6
第一阀门(降低两个半音)	○	●	●	○	○	●	●
第二阀门(降低一个半音)	●	○	●	○	●	○	●
第三阀门(降低三个半音)	○	○	○	●	●	●	●

图 4.28 铜管乐器用于演奏 7 个半音的阀门基本组合和指法（黑色圆表示按下阀门，白色圆表示抬起阀门。对小号而言，第一个阀门离吹嘴最近，第二个阀门位于中央，第三个阀门离喇叭口最近）

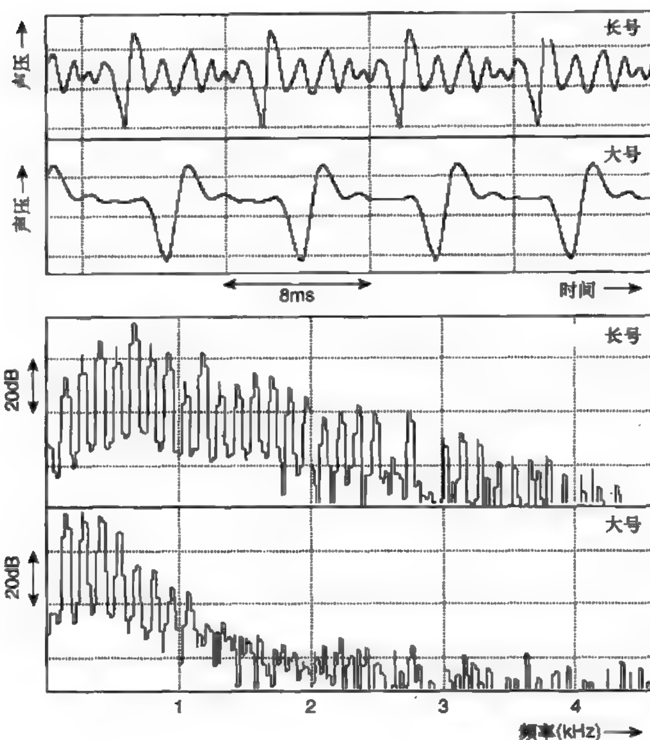
假设所采用的是平均律调音系统，在第 3 章中已经指出，相差一个半音（1/12 倍频程）的频率比是：

$$r = \sqrt[12]{2} = 1.059\ 5$$

因此，为了使音调降低 1 个半音，频率需要降低 5.95%，这也正是铜管乐器第二个阀门需要增加的管长百分数。由于降低 2 个半音的频率比是降低 1 个半音的频率比的平方（$1.059\ 5^2$=1.122 5），所以按下第一个阀门需要降低的频率或增加的管长的百分数是 12.25%。同时按下第一个阀门和第二个阀门时增加的管长为 18.2%(12.25% + 5.95%)，这对降低 3 个半音是不够的，因为降低 3 个半音需要管长增加 18.9%($1.059\ 5^3$=1.189 3)，因此演奏者还需要控制唇部来降低音调。第三个阀门在名义上是用来降低 3 个半音的，但由于降低 3 个半音需要增加的管长数量较大，因此，实际上它用于和第一个阀门组合完成降低 5 个半音的演奏。降低 5 个半音对应于管长增加 33.51%($1.059\ 5^5$=1.335 1)，减去第一个阀门增加的百分数后得到第三个阀门需要增加的管长为 21.26%(33.51%−12.25%)。这个数值要比第一个阀门和第二个阀门组合产生的 18.2% 以及降低 3 个半音所需的准确数值 18.9% 大得多。在实际应用时，以小号为例，第三个阀门通常配合伸缩管进行操作，使演奏者能够根据需要调整所增加的管长。但对长号演奏者而言不存在这样的问题，他可以通过伸缩管准确地控制增加的管长，以获得精确的音程关系。

图 4.29 所示为长号和大号演奏的 C3 波形和频谱。长号频谱的高次谐波成分比大号更明显地向高频端扩展，这个现象可以从它们的波形看出。长号的波形在一个周期内表现出比大号更多的起伏。比较两个频谱图可以看到，大号的前 3 个谐波较为突出，前 8 个谐波成分可以明显地辨认出来，而长号的第 5 次谐波占主导地位，前 29 个谐波都可以辨认出来。

图 4.29　长号和大号演奏的 C3 波形和频谱



4.4　打击乐器

打击乐器家族是乐器的一个重要组成部分，其声学特性也可以用“黑盒”模型来描述。人类经常通过敲击物体发出有节奏的声音，来吸引别人的注意力或愉悦自己。节奏在不同文化中都是构成音乐的基础，而打击乐器主要是用来产生节奏的。更多这方面资料请参看参考文献 Benade（1976）、Rossing（1990）、Hall（1991）以及 Fletcher 和 Rossing（1998）。

4.4.1　打击乐器的声源特性

打击乐器的声源通常与击打有关，最常见的方法是用棍棒、木制或金属制槌子击打物体发出声音，槌头经常用塑料膜或布包裹。只有钹是

一种例外情况，它是利用两个钹相互撞击发出声音的。这类声源被称为“脉冲”声。脉冲是非周期性信号（即波形不重复出现），其频谱是连续谱，包含所有频率成分。因此，任何打击乐器的声源都具有振动持续时间较短、包含所有频率成分的特点。所有乐器支持的振动模式都会被激发出来，每个模式都具有图 4.21 所示的拨动簧片的振动特性，即模式的振动频带越窄，其振动持续时间越长。（这种振动方式可以用小孩荡秋千作比喻。家长只要推一下，小孩和秋千就会以某一自然频率前后摆动，摆动的幅度逐渐减小为零。摆动位移随时间的变化曲线与图 4.21 所示的硬簧振动的时间响应相似。）

4.4.2 打击乐器的声音修饰器

打击乐器的声学特征主要表现在所支持的各个振动模式以及击打位置与各模式的波节和波腹位置的相互关系（参看本章前面对拨弦和击弦的介绍）。打击乐器分为 3 类：棒振动乐器（如木琴、钟琴、金属键琴和三角铁等）、膜振动乐器（鼓）和板振动乐器（钹）。无论哪种类型的打击乐器，其自然振动模式频率之间不成谐波关系。但是，棒的纵振动模式是例外情况。棒的纵振动一般用涂敷松香的布料或手套击打产生。如果是两端自由的棒，其模式频率由式（1.20）决定；如果是一端固定另一端自由的棒，则其模式频率由式（1.21）决定。

当像木琴、三角铁那样击打棒时，产生的振动模式属于横振动模式。这些模式频率之间不成谐波关系。横振动的高次模式频率与第一次模式频率之间的关系如下（Fletcher 和 Rossing 1999）：

对于自由放置在某个支点上的棒（如钟琴、木琴）：

$$f_n = 0.110\,30(2n+1)^2 f_1 \qquad (4.12)$$

其中，n 为从 2 开始的模式次数（2，3，4⋯），f_1 为第一次模式频率。

对于一端固定的棒（如金属键琴）：

$$f_2 = 0.701\,44(2.988)^2 f_1$$

$$f_n = 0.701\,44(2n+1)^2 f_1 \qquad (4.13)$$

其中，n 为从 3 开始的模式次数（3，4，5⋯），f_1 为第一次模式频率。

棒的横振动模式频率与棒长的平方成反比：

$$f_{transverse} \propto \left[\frac{1}{L^2}\right] \quad (4.14)$$

棒的纵振动模式频率与棒长成反比（见式（1.20）和式（1.21））：

$$f_{longitudinal} \propto \left[\frac{1}{L}\right] \quad (4.15)$$

因此，棒长减小一半会使横振动模式频率提高 4 倍或 2 个倍频程，而使纵振动模式频率只提高 2 倍或 1 个倍频程。除了一端固定的棒的第二次模式外（其频率是基频的 2.988 倍，非常接近于 3 倍），横振动模式的频率随着模式次数的平方而变化。表 4.1 列出了置于支点的棒（式（4.12））和一端固定的棒（式（4.13））的前 5 个模式与第一次模式的频率比以及以平均律的半音数表示的音程（附录 2 给出了频率比与半音数的转换公式）。可见，没有一个高次模式与基频的音程为半音的整数倍，都不能很好地构成音乐的音阶。各次模式频率间隔比谐波频率间隔大得多，因为它们是按照奇数平方的比例增加的（如 32，52，72，92…）。各次模式的相对强度部分地与棒的击打位置有关。

表 4.1　置于支点的棒和一端固定的棒前 5 个模式与第一次模式的频率比（式（4.12）和式（4.13））以及用半音数表示的音程（见附录 2）

棒的横振动模式	置于支点的棒		一端固定的棒	
	频率比	半音数	频率比	半音数
1（相对第一次模式）	1.000	0.00	1.000	0.00
2（相对第一次模式）	2.758	17.56	6.267	31.77
3（相对第一次模式）	5.405	29.21	17.536	49.58
4（相对第一次模式）	8.934	37.91	34.371	61.23
5（相对第一次模式）	13.346	44.86	56.817	69.93

Benade（1976）指出，实际测得的棒振动乐器的振动模式频率会与表 4.1 所示的理论值存在一些差异，这种现象是由“棒上的安装孔和为了调音将棒底部中央磨去一部分”引起的。

为了使声音具有明确的音调感，在一些棒振动乐器如木琴、马林巴琴和电颤琴（演奏音域分别为 C5（523Hz）~C8（4 186Hz）、A2（110Hz）~C7（2 093Hz）和 F3（175Hz）~F6（1 397Hz））的制作中，通常将音棒的底部做成弧形，使振动模式的频率接近于第一次振动的谐波频率。在马林

巴琴和电颤琴中，第二次模式通常调整到比第一次模式高两个倍频程；而在木琴中，常调整到比第一次模式高 1/12。这些乐器都有共鸣器，它们往往是一端封闭的管，安装在每根棒的下方。这些共鸣器的第一次振动模式调整到与棒的基频 f_0 相同，以便提高声音的响度。因此，共鸣管的长度应是 f_0 相应波长的 1/4（见式 1.21）。

对膜振动和板振动打击乐器而言，其所支持的振动模式频率主要由膜和板决定。鼓膜和钹都是圆形的，用位移波节和波腹表示的前 10 个振动模式如图 4.30 的上图所示。图中显示的是位移节点位置，它们一般呈圆形和径向轨迹。节圆和节径的个数用括号里的数字表示为：（节径数、节圆数）。鼓膜至少存在一个位移节圆，那就是被固定的膜的边缘。

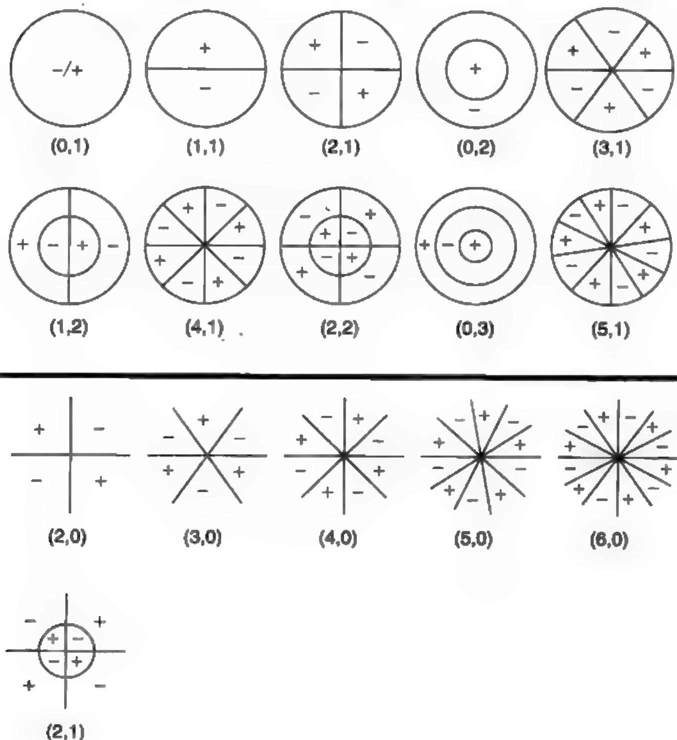


图 4.30　张紧鼓膜的前 10 个振动模式（上图）和钹的前 6 个模式（下图）（模式次数用括符表示为（节径数、节圆数），“+”、“-”表示每个模式中不同区域的振动相位关系。数据来自 Fletcher 和 Rossing 1991）

各次模式频率可以通过数学公式计算得到，但结果要比棒振动复杂得多。表 4.2 列出了各次模式与第一次模式的频率比（Fletcher 和 Rossing 1999）以及用半音数（用附录 2 给出的公式换算）表示的音程差。和棒

振动的情况相同，没有一个模式频率与第一次模式成整数倍半音关系或频率比为整数比。因此，这些振动模式之间不成谐波关系。

表 4.2 不带共鸣腔的圆膜（左）、像定音鼓一样带共鸣腔的圆膜（中）和理想圆形板（右）的前 10 个模式、与第一次模式的频率比（Fletcher 和 Rossing 1999）以及用半音数表示的音程差（频率比和半音数的变换公式参看附录 2）（星号表示在定音鼓通常击打位置击打时不被明显激发的模式）

圆膜			圆板	
模式	理想圆膜	定音鼓圆膜	模式	钹
	频率比（半音数）	频率比（半音数）		频率比（半音数）
(0,1)	1.000(0.0)	1.70 (9.19)*	(2,0)	1.00(0.0)
(1,1)	1.59(8.1)	2.00(12.0)	(3,0)	1.94(11.5)
(2,1)	2.14 (13.1)	3.00 (19.0)	(4,0)	3.42(21.3)
(0,2)	2.30(14.4)	3.36(21.0)*	(5,0)	5.08(28.1)
(3,1)	2.65(16.9)	4.00(24.0)	(6,0)	6.90(22.4)
(1,2)	2.92(18.6)	4.18(24.7)*	(2,1)	8.63(37.3)
(4,1)	3.16(19.9)	4.98(27.8)	-	-
(2,2)	3.50(21.7)	5.34(29.0)*	-	-
(0,3)	3.60(22.2)	5.59(29.8)*	-	-
(5,1)	3.65(22.4)	5.96(30.9)	-	-

鼓主要由鼓膜或称为“鼓头”构成，鼓膜一般由合成材料或动物的皮制成，并张紧在一个圆形的支架上。在早期音乐中使用的像小鼓这样的小型手鼓一般由一个圆柱形环组成，鼓膜张紧在其中一端，另一端则呈开放状态。铃鼓具有类似的结构，但是在圆柱形环上成对安装有类似于钹的金属圆片或铃铛。当用指关节或指尖敲打铃鼓时，铃铛会同时发出清脆短促的铃声；而当用拇指沿着铃鼓边缘摩擦运动时，铃铛会持续发出声音。体型较大一些的鼓还有康茄鼓和小手鼓。它们呈长圆柱形，一般也只有单个鼓头，通常用演奏者的手进行敲打。

套鼓主要由大鼓、响弦鼓（小军鼓）和两个或两个以上不同尺寸的长鼓组成。包括管弦乐队的大鼓在内，这些鼓都有两个鼓头，分别安装在圆柱形鼓体的两端。大鼓和响弦鼓属于无调打击乐器。通常将两端鼓膜调整为具有不同张力，使非谐波关系的各模式频率在较宽的频率范围分布。当它们被击打发声时，其声音无特定音调感，因此可以用来为各种调性的音乐提供节奏。响弦鼓（在进行曲乐队中称为“边鼓”）的底部

鼓头安装有若干金属弦，横穿过鼓膜。当其上部鼓头被打击时，金属弦会相对于鼓膜产生振动，发出独特的声音。金属弦还可以用工具从鼓头松开，使鼓发出没有响弦的声音。另一方面，每个长鼓的上鼓头和下鼓头通常被调音到音高相一致，使发出的声音有更明确的音高感。套鼓经常以向上或更经常地向下滚奏的方式用于旋律的间隙。

虽然长鼓能产生较为明确的音调感，但它不是音调感最为明确的鼓。管弦乐队中主要使用的是壶形鼓。目前管弦乐队一般使用 2 ~ 5 个壶形鼓，由一个演奏者进行演奏。这样的一组壶形鼓称为定音鼓。由于定音鼓能够方便快速地改变音调。因此当音乐发生转调时，作曲家可以在新的调式里继续使用定音鼓。每个定音鼓是由一张鼓膜张紧在一个金属碗上构成，放置在一个支架上。定音鼓的绝对或参考音高的调整是由鼓边缘上的螺丝（通常有 8 个）完成的，以获得合适的绝对音调。在演奏过程中，通过踏板操纵的杠杆张力调整系统，可以使音调以半音的步长变化，其变化范围一般是 5 个半音。当杠杆从一个音移动到另一个音时，鼓膜的张力随之变化。当正好在移动杠杆之前击打鼓面时，就会产生一种称为滑奏的向上或向下的变化音调。这种演奏方法有时用于产生某种特殊的音乐效果。

定音鼓的振动模式与图 4.30 相同，但是，由于膜后空腔的作用以及所采用的击打位置不同，其振动模式频率与背后无空腔的膜（表 4.2 中的"理想膜"）有很大不同。Hall（1991）对此描述如下："当鼓面的一些区域相对于碗状空腔向里和向外移动时，另一些区域则向相反的方向移动（分别向外和向里）。"这些反相振动的区域在图 4.30 中分别用"＋"和"－"表示。（0，1）模式是个例外，其整个膜片整体地向里或向外移动，使碗状空腔内的空气压缩或膨胀，用符号"+/－"表示。在实际应用中，往往在腔体上开有一个小孔，使空气对应模式的运动状态流出或进入小孔，从而消耗一部分能量，对该模式的振动起到一定的抑制作用。这个小孔对其他振动模式的影响不大，因为其他振动模式都存在互补的振动状态区域，在图 4.30 中用等量的"＋"号和"－"号表示。Hall 指出，定音鼓的击打位置应位于从鼓膜中心到鼓膜边缘的 1/2 ~ 3/4 距离处，这个位置与具有两个或三个节圆的振动模式，如图 4.30 所示的（0，2）、（1，2）、（2，2）和（0，3）模式的节圆位置较为接近。由于击打会在这些模式的位移波节位置产生较大的振动位移，因此，这些振动模式很难显著地被

激发出来（参看 4.2.1 节有关拨弦时不被激发的模式的阐述）。

总之，当定音鼓的击打位置位于从鼓膜中心到鼓膜边缘的 1/2 ~ 3/4 距离处时，只有仅有一个节圆（指每个模式都有的位于边缘的节圆）的振动模式对声音有贡献，其中第一次振动模式（0，1）除外。由于碗状腔体内空气的作用，使鼓膜的有效质量增大，因此所激起的各次振动模式频率有所降低，频率降低幅度与空腔的形状有关。定音鼓制作者最终要达到的目的是使激起的各次振动模式频率接近于同一系列的某几个谐波频率。表 4.2 列出了理想状态的圆膜（振动模式如图 4.30 所示）和定音鼓圆膜的最低 10 个模式的频率比（Rossing 1989），那些对定音鼓的声音贡献不大的模式如（0，1）、（0，2）、（1，2）、（2，2）和（0，3）用星号注明。可见，剩余的各次模式频率非常接近于二次、三次、四次、五次和六次谐波频率，使得定音鼓能够成为管弦乐队的乐器之一。只是其声音缺失基频，但这并不影响听觉对其基频音调的感觉，这方面内容可参看本书 3.2.1 节。因此，只要选择合适的击打位置，使表 4.2 中那些带星号的模式被有效抑制，定音鼓就能发出具有很强音调感的声音。

由此可知，演奏者确实能够通过选择击打位置控制定音鼓的输出频谱。如果选择中心位置击打，则发出的声音并不令人满意。因为几乎所有的模式在中心位置都存在一个节点，因此选择中心位置击打时大多数模式不会被激发出来。另一种控制声音的方式是使用不同的槌头。小而坚硬的槌头激振强度大而范围小，因此会激发出更丰富的高频模式；而大而柔软的槌头激振强度较小而范围较大，因此更容易激发起低频振动模式，其声音显得沉闷、暗淡。早期常采用布包鼓头来产生发闷的声音效果。管弦乐队中的定音鼓通常采用滚奏和反复的演奏技巧。

图 4.30 的下半部分表示钹的前 6 个振动模式（Rossing 1989）。值得注意的是，与张紧的圆膜不同，板的边缘不需要支撑，因此其低频模式不存在节圆。Rossing 指出，高于 6 次的模式往往是由两个以上的振动模式组成，其振动模式非常复杂，因此难以用图形表示出来，但是，它们确实存在至少一个节圆。由于各模式的振动频率并不与谐波频率相近，因此钹的声音没有明显的音调感。钹可以用不同的方式激振，例如，管弦乐队中常用两只钹相对撞击发出声音，或套鼓中用坚

硬或柔软的鼓槌敲击发出声音等。由于所有模式都在圆盘的边缘产生波腹，因此在此部位撞击或敲击能激发出所有的振动模式。和定音鼓一样，演奏者可以通过击打工具的使用在一定程度上控制输出声音的频谱。

4.5 语声和歌声

歌唱可能是所有乐器中最具表现力的乐器。任何一个会说话的人都会唱歌，但并不是所有人都能成为歌唱家或流行音乐歌星。当人们对歌唱教师的工作和他们如何取得这些成果充满了神秘感时，出于某种自身的需求，歌声的声学特性也成为了新的研究课题，主要包括以下几个方面：

（1）成人男声与女声的区别；

（2）歌唱训练的效果；

（3）儿童听音技能的开发；

（4）“音盲”；

（5）不同歌唱风格的声学特点；

（6）不同声学特性对合唱效果的影响；

（7）歌声的电子合成；

（8）合唱中的歌声融合；

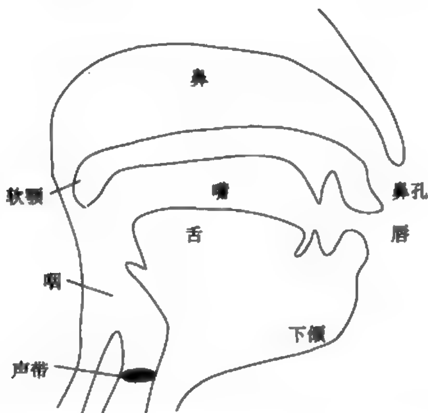
（9）独唱的发声原理。

人类能够很好地发出他们想要的声音，因此，在音乐技术中，对语声和歌声发声机理的了解是声音合成技术的基础，它能够帮助有效地制定声音合成的方案。本节主要利用输入—系统—输出模型介绍人类歌声的声学特性，同时指出语声和歌声的主要区别。本节只是对这些内容进行简述。关于语声和歌声声学特性的详细介绍可参看其他文献资料（如 Fant 1960、Fry 1979、Borden 和 Harris 1980、Baken 1987、Baken 和 Danilof 1991、Kent 和 Read 1992、Howard 1998、Howard 和 Angus 1998、Benade 1976、Sundberg 1987、Bunch 1993、Bejonckere 等 1995、Howard 1999）。

4.5.1　歌声的声源特性

歌声的声源是位于喉部的声带，声带在来自肺部的气流作用下保持振动。歌声的声音修饰器是从喉部到唇部和鼻孔的空腔，称为“声道”。通过移动“发音器官”，例如下颌、舌、唇等（见图 4.31），可以改变声道的形状和大小。当我们说话或歌唱时，声道的形状不断发生变化而发出不同的音节。软颚就像阀门一样可以打开或关闭气流通向鼻腔（鼻）的通道。

图 4.31　声道截面图



当唱出某个音符时，正常的喉部声带（由两个部分组成）会有规律地关闭和打开，产生周期性的振动。因此，女高音 A4（f_0=440Hz）的声带每秒会完成 440 次完整的关闭和打开周期。歌唱家可以采用两种方式改变声带振动基频 f_0：一是通过改变声带肌肉组织的张力来改变声带本身的硬度，二是通过使两侧声带的一部分保持静止状态来改变声带的振动质量。声带自身的这种调节能力使许多训练有素的歌唱家能够达到两个八度以上的演唱音域。

声带振动也是“伯努利效应”的结果，和铜管乐器演奏者双唇的振动方式极为相似，但声带振动具有关闭比打开快得多的特点。就像拍手一样，每当声带关闭的瞬间，就会产生一个声压脉冲。当唱出某个音符时，声带会有规律地打开和关闭，形成一系列周期性的声压脉冲信号（见图 4.32），作为声道的输入信号。声音的音调则是由每秒的脉冲个数决定的。

图 4.32 中显示的声压脉冲为负脉冲，这是因为当声带快速闭合时引起来自肺部的气流中断，使声带前的声压迅速降低。基频周期就是相邻两个脉冲的时间间隔。Benade(1976) 指出，虽然铜管乐器演奏者的唇部振动与说话者和歌唱者的声带振动相似，但二者并不完全相同。声带的振动基本上不受声道的影响，而铜管乐器演奏者唇部的振动却与音管有密切关系。

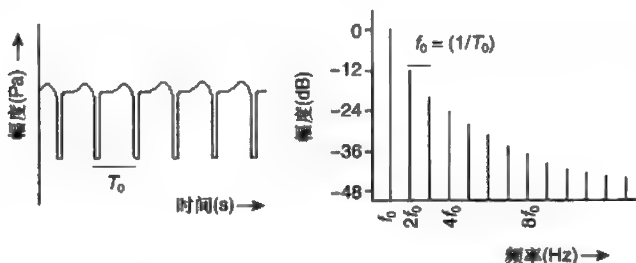


图 4.32　正常声带振动产生的理想化声压波形（左）和频谱（右）（T_0 表示声压波形周期，f_0 表示基频）

图 4.33 为以穿过声带间隙的理想气流表示的声带振动过程的图解。由于两侧声带之间的间隙称为“声门”，因此这种气流称为“声门气流”。振动周期中的 3 个主要阶段一般是指：已闭合阶段（声带完全闭合）、打开阶段（声带逐渐打开）和关闭阶段（声带关闭过程）。打开和关闭阶段一并在图中用打开的阶段表示，是指有一定气流流过的时间段。值得注意的是，由于声带结构的原因，在已闭合阶段声带不可能沿着长度方向完全处于闭合状态，因此，在已闭合阶段气流并不一定是零（Sundberg 1987，Howard 1998、1999）。

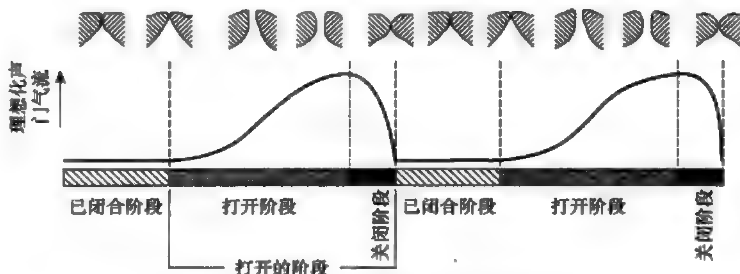


图 4.33　以声门气流表示的声带振动过程的两个周期的图解（图中表示出声带振动的 3 个阶段）

无论是演讲、戏剧表演还是歌唱，声带振动的特性随着发声训练的不同而不同。在每个振动周期中，两侧声带相接触的时间称为“喉闭合商（larynx closed quotient）”或“CQ”。研究表明，CQ 与成年男、女歌唱家是否能够较有效地发出声音有关（Howard 等 1990，Howard

1995）。通过对训练过的和没有训练过的歌唱者的测试发现，训练过的男歌手在所有音域具有较高的 CQ 值，对训练过的女歌手而言，CQ 值随着音调的升高以一定的形式增加。Howard 等指出，较高的 CQ 值能够更有效地发出声音是基于以下 3 个原因：（1）每个周期中，声能可能通过声门到达肺部被完全吸收的时间减少了；（2）由于每个周期中较少的声能通过声门损失掉，因此发出的声音能够持续较长时间；（3）由于每个周期中较少的气流通过声门，因此减小了呼吸声，改善了音质。

在说话或歌唱时，由声带振动产生的周期性声压脉冲的频谱包含所有谐波成分，各谐波成分大小按频率每增大一个倍频程下降 12dB 的规律变化（见图 4.32 的右图），因此，频率每增大一倍，即提高一个八度，幅度减小 12dB。图中第 1 次、第 2 次、第 4 次和第 8 次谐波的幅度反映了这个特性。

歌唱时声源的频谱形状基本上保持不变，但是，各谐波成分的大小随着频率增大以每倍频程 -12dB 下降的斜率会随着艺术效果、演唱风格的不同而改变，专业歌唱家往往以此进行声音设计（Sundberg 1987）。当演唱不同音高的音符时，谐波成分之间的频率间隔会随之改变。训练有素的歌唱家，特别是那些采用西洋歌剧唱法的歌唱，会产生“颤音”的效果，即歌唱时基频 f_0 以大约 5.5Hz ~ 7.5Hz 的速率变化，波动范围为 ±0.5 ~ ±2 个半音（Dejonckere 等 1995）。

4.5.2　歌声的声音修饰器

声道（见图 4.31）的声学特性会对声带产生的周期性脉冲起修饰作用。用声学术语来说，声道可以看成一个一端封闭的管（喉端类似于“气流控制簧”，视为封闭端，唇部视为开放端）。成人男性的声道长度大约为 17.5cm。当英文单词“announcer”的最后一个元音发出时，声道的状态称为“中性”。此时各发音器官处于放松状态，软颚（见图 4.31）抬高阻断通往鼻腔的气流，这类元音也称为“非鼻音化”元音。中性状态的声道非常接近于截面均匀的长管，因此，可以用圆柱形封闭管的公式计算这个元音的声道驻波模式频率。

例 4.3　试计算中性成人男性声道的前 3 个模式频率（取声速为 344m/s）。

声道长度为 17.5cm 或 0.175m。

根据式（4.9），基频或第 1 次模式频率为：

$$F_{\text{stopped}(1)}=\left[\frac{c}{4L_s}\right]=\left[\frac{344}{4\times0.175}\right]=491.4\text{Hz}$$

由式（4.10）得，较高模式的频率为：

$$f_{\text{stopped}(n)}=(2n-1)f_{\text{stlooed}(1)}$$

其中，n=1，2，3，4…

因此，第 2 次模式频率（取 n=2）为：3×491.4=1 474Hz

第 3 次模式频率（取 n=3）为：5×491.4=2 457Hz

例 4.3 得出了中性元音的模式频率。为了方便使用，这 3 个频率常用 500Hz、1 500Hz 和 2 500Hz 来近似。在语声和歌声的声学研究中，这些模式频率常被称为“共振峰”。图 4.34 所示为发出单词“fast”、“feed”和“food”中元音时声道的理想化频率特性曲线。图中，每个共振峰的中心频率按从低到高分别用“F1”或“第 1 共振峰”、“F2”（第 2 共振峰）和“F3”（第 3 共振峰）表示。共振峰是声道本身的共振频率，并随着声道尺寸和形状的改变而改变。当人们说话或歌唱时，发音器官的位置不断变化，从而使共振峰随之变化。在分析语声中的不同语音时，往往只考虑第 1、第 2 和第 3 共振峰，这是因为对不同语音而言，只有前 3 个共振峰有明显的区别。在实验室中，前 6 ~ 7 个共振峰能够被区别开来，但一般认为更高的共振峰只对每个人说话或歌唱的音色有影响。然而，第 3 次以上的共振峰确实对歌声的表现力和风格起非常重要的作用。

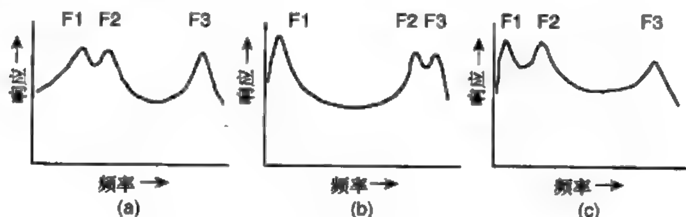


图 4.34　单词“fast”、“feed”和“food”中元音对应的声道理想化频率特性曲线

为了发出不同的语音，声道的形状可以通过发音器官位置的变化来改变，从而改变声道的声学特性。在有簧木管乐器一节中介绍的关于局部缩窄或扩大的理论也可以在此得到应用（参看图 4.23）（Kent 和 Read 1992）。图 4.35 所示为中性非鼻音化元音声道的前 3 个共振峰的位移波节和波腹位置，可用图 4.23 的右上图进行对比。回顾书中与图 4.23 相关的文字叙述，采用相同的分析方法，可以对声道的局部缩窄（或局部放大）对前 3 个共振峰频率的影响进行预测，如图 4.36 所示。由于所有的共振峰在唇部都存在一个体积速度（位移）波腹，因此，在唇部的缩窄会使所有共振峰频率下降。（需要指出的是，另外还有两种降低所有共振峰频率的方法，即通过向前延伸唇部和降低喉部的方法加长声道，从而降低共振峰频率。）

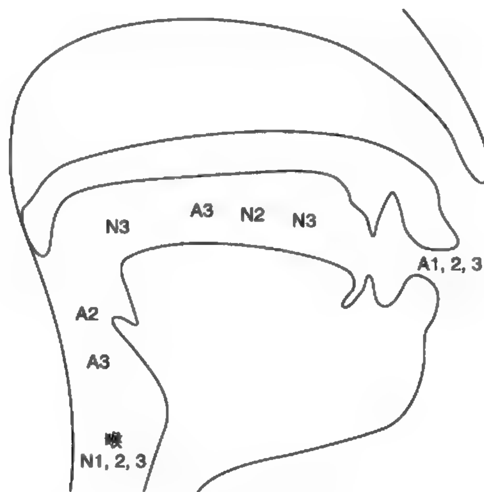


图 4.35　中性非鼻音化元音声道的前 3 个模式（或共振峰 F1、F2、F3）的位移波节和波腹位置

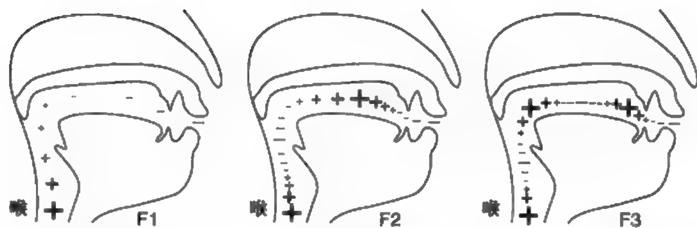


图 4.36　声道的局部缩窄对共振峰频率的影响

表 4.3 列出了由 Peterson 和 Barney（1952）统计的一些元音的男人、女人和儿童的常用共振峰频率。根据发音器官的位置，可以对这些元音

的共振峰频率进行估计。例如，“beat”中的元音在声道的舌前部有一个收窄，大概处于 N2 和 N3 的位置（见图 4.35），由图 4.36 可知，与中性声道相比，F1 频率有所降低，而 F2、F3 则有所提高；再如，“part”中的元音在声道的 A2 和 A3（见图 4.35）处有一个收窄，结果使 F1 较中性频率有所提高，而 F2 和 F3 较中性频率有所降低。“boot”中的元音在唇部有一个收窄，同时唇部形成圆形并向前有一定的延伸，因此所有的共振峰比中性都有所降低。这些共振峰变化可以从表 4.3 得到验证。

表 4.3　一些元音的男人、女人和儿童的平均共振峰频率（Hz）（Peterson 和 Barney 1952）

下列单词中的音	男人			女人			儿童		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
Beat	270	2 300	3 000	300	2 800	3 300	370	3 200	3 700
bit	400	2 000	2 550	430	2 500	3 100	530	2 750	3 600
bet	530	1 850	2 500	600	2 350	3 000	700	2 600	3 550
bat	660	1 700	2 400	860	2 050	2 850	1 000	2 300	3 300
part	730	1 100	2 450	850	1 200	2 800	1 030	1 350	3 200
pot	570	850	2 400	590	900	2 700	680	1 050	3 200
boot	440	1 000	2 250	470	1 150	2 700	560	1 400	3 300
book	300	850	2 250	370	950	2 650	430	1 150	3 250
but	640	1 200	2 400	760	1 400	2 800	850	1 600	3 350
pert	490	1 350	1 700	500	1 650	1 950	560	1 650	2 150

歌声的“输入—系统—输出”模型是由声带振动产生的声激励（输入）、声道频率特性（系统）和输出组成。这个过程一般可以从频谱的角度来考虑。在歌唱时，输入和声道频率特性都随时间而变化。图 4.37 所示为“fast”中元音的 3 个不同音高的“输入—系统—输出”模型，用于说明不同音高演唱的声学效应。

输入分别是来自声带振动的声激励频谱（参看图 4.32），输出则是声道的“fast”元音的频率响应作用于输入的结果。这种作用相当于将输入中的各次谐波大小分别与声道响应在相应频率的大小相乘，从而使声道的共振峰结构在输出频谱中表现出来。在这个图例中，声道响应有 3 个共振峰，由图可知，对于较低的两个音，声道的共振峰结构可以明显地在输出频谱中看到；而对于最高的那个音，由于谐波成分的间隔过大，使得声道的共振峰结构很难在输出频谱中看到。

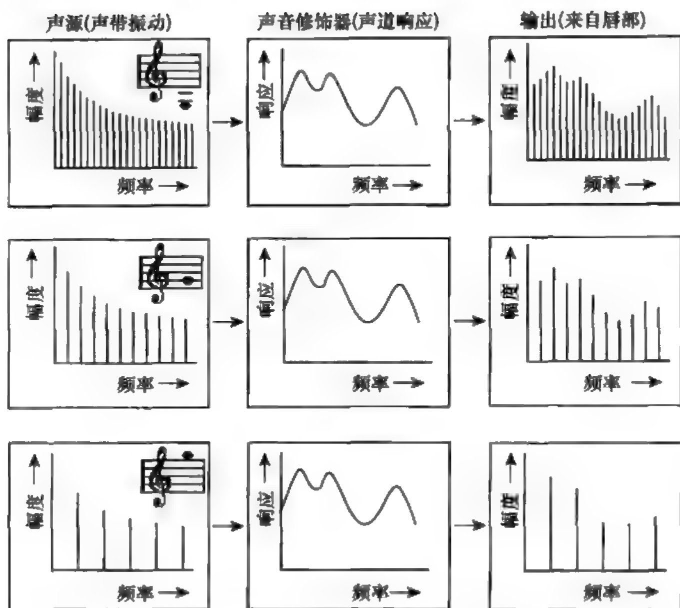


图 4.37 "fast" 中元音以八度间隔的 3 个不同音高演唱的理想化"输入—系统—输出"模型

输出频谱中的共振峰结构对聆听者辨别不同的元音是非常重要的。由图 4.37 可知，当音高在中央 C 之上的 G 和高八度 G 之间时，元音的辨别随着音调的提高变得越来越困难。这一点可以通过让女高音按图中所示的从中音 G 到高音 G 演唱不同的元音，然后聆听来进行测试。在实际演唱时，专业女高音往往通过控制声道的形状，使较低的共振峰频率与一些谐波频率相重合，此时声带通过声道发出的音量最大。通过这种方法，女高音可以发出高声强的声音，使歌声更富于表现力。这种歌唱技巧一般在中央 C 以上的 G 开始使用，将声道共振峰"调准（tuned in）"到每个音符的基频，付出的代价是语音的清晰度降低了。

男高音则不需要运用调准效应来提高音量，因为声道共振峰频率与男高音音域 f_0 的比值要比女高音的大。然而，所有不使用扩音器的歌唱者都需要把声音突出于伴奏或伴唱之上，特别是当整个交响乐队为其伴奏或在一个较大的厅堂进行演唱时。图 4.38 能够说明歌唱家用什么方法达到这个目的，它表示了以下三种情况的理想化输出频谱：

（1）专业歌唱家读出歌剧咏叹调歌词时；

（2）交响乐队演奏咏叹调的伴奏时；

（3）歌唱家演唱带交响乐伴奏的咏叹调时。

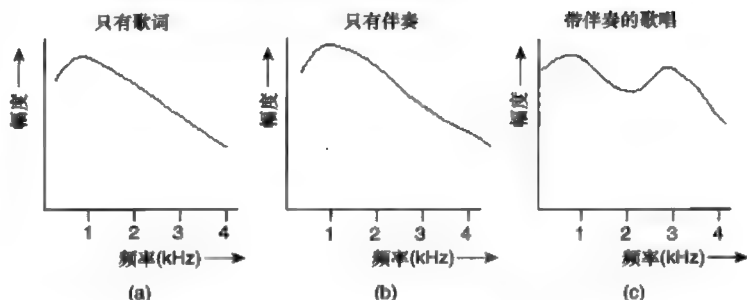


图 4.38　理想化输出频谱（Sundberg 1987）（a）歌唱家读出歌剧咏叹调歌词（b）交响乐队演奏咏叹调的伴奏（c）歌唱家演唱带交响乐伴奏的咏叹调

需要指出的是，图中（a）和（b）的幅度不能直接进行比较（即图中曲线并不表明朗读声和交响乐伴奏声一样响），因为为了便于对比，图中给出的是相对幅度值。

由图可知，只有朗读和只有伴奏的理想化频谱大致上呈现相同的频谱形状。当歌唱家带伴奏进行演唱时，其频谱在低频段具有与只有朗读和只有伴奏的频谱基本相同的形状，但是，在大约 2.5kHz ~ 4kHz 之间出现了一个宽的峰值，这个峰值只与歌唱家歌唱时的声输出有关，而与朗读和伴奏无关，因为这个峰值既不存在于只有朗读的频谱中，也不存在于只有伴奏的频谱中。这个频谱峰值与声道响应的共振峰很类似，因此称为“歌唱者共振峰（singer’s formant）”。由于在这个共振峰频带，歌声的输出远大于伴奏，因此，歌唱家的声音能在伴奏交响乐之上被听到。这就是专业歌唱演员的歌声特别“响亮”的原因，它是通过降低喉部和扩大咽部（见图 4.31）获得，这种演唱技巧一般为西洋歌剧演唱者所采用。（图 5.5 的下图是专业男高音 CD 录音的分析，可以很明显地看到“歌唱者共振峰”。）

声乐教师从小学就开始训练学生获得这种演唱技巧，他们会对学生说：“像打哈欠时那样张开嘴唱歌”，或者说：“唱歌的时候想象你刚刚咽下一个苹果，并且苹果卡在你的喉部”。Sundberg（1987）在研究“歌唱者共振峰”与发音器官之间的关系时做出以下阐述：“它与喉管有密切关系”，并且得出下述结论：“咽管必须拉长，而且，咽部在喉管开口位置的截面积必须比此开口面积大 6 倍以上”。

专业歌唱有一个复杂的任务，就是极大扩展作为语声乐器的发声器官的功能。但是，人们更应该认识到，声带的基本功能更多的是作为阀

门保护肺部，而不是通过说话或歌唱作为人类交流的基本工具。

参考文献

Askenfelt, A. (ed.) (1990). Five Lectures on the Acoustics of the Piano, with compact disc, Publication No. 64. Stockholm: Royal Swedish Academy of Music.

Audsley, G.A. (1965). The Art of Organ-Building, 2 Vols. New York: Dover (reprint of 1905 edition, New York: Dodd, Mead and Company).

Backus, J. (1977). The Acoustical Foundations of Music. New York: Norton.

Baken, R.J. (1987). Clinical Measurement of Speech and Voice. London: Taylor and Francis.

Baken, R.J. and Danilof, R.G. (1991). Readings in Clinical Spectrography of Speech. San Diego: Singular Publishing Group.

Benade, A.H. (1976). Fundamentals of Musical Acoustics. New York: Oxford University Press.

Borden, G.J. and Harris, K.S. (1980). Speech Science Primer. Baltimore: Williams and Wilkins.

Bunch, M. (1993). Dynamics of the Singing Voice. New York: Springer-Verlag.

Campbell, M. and Greated, C. (1998). The Musician’s Guide to Acoustics. Oxford: Oxford University Press.

Dejonckere, P.H., Hirano, M. and Sundberg, J. (eds) (1995). Vibrato. San Diego: Singular Publishing Group.

Fant, C.G.M. (1960). Acoustic Theory of Speech Production. The Hague: Mouton.

Fletcher, N.H. and Rossing, T.D. (1999). The Physics of Musical Instruments (2nd edn). New York: Springer-Verlag.

Fry, D.B. (1979). The Physical of Speech. Cambridge: Cambridge University Press.

Hall, D.E. (1991). Musical Acoustics: An Introduction (2nd edn). Belmont, CA:

Wadsworth Publishing Company.

Howard, D.M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. Journal of Voice, 9, 163–172.

Howard, D.M. (1998). Practical voice measurement, in: The Voice Clinic.

Handbook (T. Harris, S. Harris, J.S. Rubin and D.M. Howard, eds). London: Whurr Publishing Company.

Howard, D.M. (1999). The human singing voice, in: Killers in the Brain (P. Day, ed.). Oxford: Oxford University Press, pp. 113–134.

Howard, D.M. and Angus, J.A.S. (1998). Introduction to human speech production, human hearing and speech analysis, in: Speech Technology for Telecommunications (F.A. Westall, R.D. Johnson and A.V. Lewis, eds). London: Chapman and Hall, pp. 30–72.

Howard, D.M., Lindsey, G.A. and Allen, B. (1990). Towards the quantification of vocal efficiency. Journal of Voice, 4, 205–21. (See also errata: (1991). Journal of Voice, 5, 93–95.)

Hurford, P. (1994). Making Music on the Organ (6th edn). Oxford: Oxford University Press.

Hutchins, C.M. (ed.) (1975a). Musical Acoustics, part I: Violin Family Components. Pennsylvania: Dowden, Hutchinson and Ross Inc.

Hutchins, C.M. (ed.) (1975b). Musical Acoustics, part II: Violin Family Functions. Pennsylvania: Dowden, Hutchinson and Ross Inc.

Hutchins, C.M. (ed.) (1978). The Physics of Music. Reprints from Scientific American, San Francisco: W.H. Freeman and Company.

Kent, R.D. and Read, C. (1992). The Acoustic Analysis of Speech. San Diego: Singular Publishing Group.

Martin, D. and Ward, D. (1961). Subjective evaluation of musical scale temperament in pianos. Journal of the Acoustical Society of America, 33, 582–585.

Nederveen, C.J. (1969). Acoustical Aspects of Woodwind Instruments. Amsterdam: Frits Knuf.

Norman, H. and Norman, H.J. (1980). The Organ Today. London: David and Charles.

Peterson, G.E. and Barney, H.E. (1952). Control methods used in the study of vowels. Journal of the Acoustical Society of America, 24, 175–184.

Proctor, D.F. (1980). Breathing, Speech and Song. New York: Springer-Verlag.

Rossing, T.D. (1989). The Science of Sound. New York: Addison-Wesley.

Sumner, W.L. (1975). The Organ (5th edn). London: Macdonald and Company.

Sundberg, J. (1987). The Science of the Singing Voice. DeKalb: Illinois University Press.

Sundberg, J. (1989). The Science of Musical Sounds. San Diego: Academic Press.

Taylor, C.A. (1976). Sounds of Music. London: Butler and Tanner Ltd.

第 5 章　音色感知与听觉幻觉效应

5.1　什么是音色

音乐家用于描述乐音的重要术语有三个，音高和响度是其中的两个，另一个就是“音色”。音高与诸如乐谱上的音符、琴键、旋律、和声、音律和节目的音调等有关，响度则与音乐的动态特性（例如表示声音强度的符号 p、pp、mp、mf、f、ff 等）、合奏音乐各成员之间（例如个人之间、合唱队与乐队之间、独唱与伴唱之间）的平衡有关。用于描述声音质量的音色术语包括：柔和的、丰富的、隐蔽的、开放的、暗淡的、明亮的、黑暗的、刺耳的、令人烦躁的、粗糙的、尖锐的、圆润低沉的、压抑的、苍白的和死气沉沉的等。因此，音色是用来描述具有特定音高和响度的声音的质量感受或声音特质的。

听音者能够对声音的音高或响度按照从“高”到“低”的顺序排列，但是，与音高和响度不同，音色的主观评价不存在一个度量尺度。美国国家标准化研究所关于音色的正式定义说明了这一点：“音色是指声音在听觉上产生的某种属性，听音者能够据此判断两个以同样方式呈现、具有相同音高和响度的声音的不同”（ANSI 1960）。换句话说，两个相同响度感觉和音高感

觉的声音是凭借它们的音色加以区别的（参考本书 CD 的第 62 条）。乐音的音色是指，当不同的乐器演奏具有相同音高、响度和持续时间的音符时，听音者用来判别所演奏乐器的声音感觉。Scholes(1970) 对音色的定义中还涉及一些音色评价术语："音色是指声音的质量——粗糙的或光滑的、响亮的或微妙而富于穿透力的、像小号那样"鲜红"、像大提琴那样"深棕"或像长笛那样"银白"。这种颜色类比会很自然地来到每个人的头脑里。……声音中一个也是唯一一个对音色有影响的因素是泛音的存在与否以及它们的相对强弱（表 3.1 列出了泛音与谐波之间的关系）"。虽然这种颜色类比不一定会自然地出现在每个人的脑海里，但 Scholes 在上述的后面部分对不同音色声学特性的阐述对乐音音色的声学探讨具有一定的贡献。

在考虑乐器音符的音色时，音色只与乐音的某些特性有关。当改变这些特性时，不会从整体上改变乐音的音高、持续时间和响度。这些与音色有关的特性主要有乐音所呈现的频率成分、在乐音持续过程中频率及其幅度随时间变化的方式等。在第 4 章中，我们从乐器声输出角度对乐器的声学特性进行了分析，这些声输出是乐器声输入经过声音修饰器作用的结果（见图 4.2）。由于建立模型具有一定的复杂性，所以在第 4 章没有分析乐音在开始阶段（从无到有）和在结束阶段（从有到无）的声学过程。为了全面分析乐音，较为简便的方法是将乐音分为 3 个阶段进行分析："开始"或"起振"（乐音从无到有的开始部分）、"稳态"（乐音的主要部分）和"结束"或"衰减"（输入能量中断后，乐音从有到无的结束部分）。乐音的开始和结束部分持续时间大约是几十毫秒（或百分之几秒），而乐音在开始和结束阶段的变化过程，特别是开始阶段的变化过程，对乐音的音色起非常重要的作用。

本章将从不同音色声音的声学特性和听觉感知声音的心理声学两个方面分析音色。最后，从管风琴如何能够合成不同音色的角度，进一步讨论管风琴的声学特性，以此增加对音色感知的认识。

5.2　音色的声学特性

在第 4 章中，乐音的声学特性大多数用乐器声输出的波形和频谱来

表示（见图 4.17、图 4.22、图 4.24 和图 4.29）。除了拨弦乐器诗琴和吉他的图（见图 4.11）以外（其波形是整个乐音的波形，频谱是某一时刻的频谱图），其他图形显示的是乐音在稳态阶段几个周期的波形，频谱则是对稳态阶段不同时间点测量的几个频谱的平均值。进行平均的频谱个数取决于乐音稳态阶段持续时间的长短。对于第 4 章所示的单音，根据音符持续时间的长短，频谱的平均是按大约每（1/4 ~ 3/4）s 取一个测量点进行的。另一种方法是根据重复周期的个数选取测量点。例如，对于 A4（f_0=440Hz），1/4s 对应于 110 个周期，3/4s 对应于 330 个周期，对于 C4（f_0=261.6），则分别对应于 66 ~ 198 个周期。这种平均频谱常用于乐音的频谱分析中，称为“长期平均频谱”或“LTAS”（long-term average spectra 的缩写）。采用 LTAS 的主要优点是能够突出乐音在稳态阶段的频谱特征，同时，通过平均可以去除在稳态过程一直存在并且不断变化的较强的背景噪声频谱。

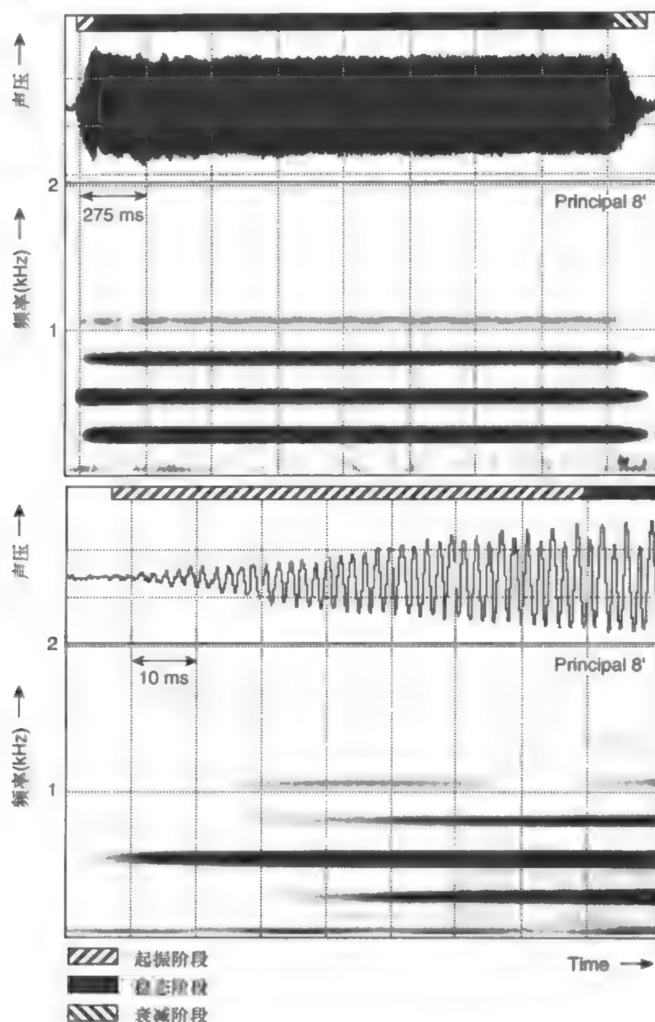
但是，LTAS 不能反映快速变化的声学特性，例如乐音的开始阶段和结束阶段的频谱，因为平均的结果会使频谱特性丢失。当研究音色时，我们不仅对开始和结束阶段的频谱感兴趣，而且对它们怎样随时间变化感兴趣。因此，需要一种表示方法能够记录乐音在整个持续过程中随时间变化的特性。一种常用于语声的分析方法是幅度、频率和时间图形，称为“声谱图（spectrogram）”。其纵轴是频率轴，横轴是时间轴，幅度的大小用灰颜色的深浅表示，或用光谱的不同颜色表示。

图 5.1 的上半部表示管风琴音栓 principal 8'（开放式哨管）C4 的波形和声谱图，同一个乐音的 LTAS 频谱见图 4.17。从图 4.17 的 LTAS 频谱看出，该音符的一次和二次谐波占主导地位，三次谐波大约比一次谐波低 8dB，四次、五次、七次和八次谐波也可分辨，但其幅度至少比一次谐波低 25dB。

声谱图显示了幅度（灰度）随频率（纵轴上的刻度）和时间（横轴上的刻度）变化的特性。因此，在声谱图中，谐波表现为水平方向的直线，其频率与纵轴上的刻度相对应，横轴的位置表示该谐波持续时间的长短，谐波幅度的大小用直线的灰度表示。图 5.1 中有 3 条清晰可见的直线，表示乐音中的前 3 个谐波（因为频率轴采用线性刻度，所以直线是等间隔的）。第 1 条和第 2 条直线比第 3 条直线更黑（更粗），反映了图 4.17 所

示的幅度差异，第 4 条、第 5 条和第 7 条直线依稀可见，它们与第 1 次谐波的幅度差异可以从线条的灰度差别看出。

图 5.1　管风琴音栓 principal 8'（开放式哨管）C4 的完整波形和声谱图（上图）以及起振阶段的波形和声谱图（下图）（图中标出了起振、稳态和衰减阶段）



5.2.1　乐音的包络

图 5.1 所示的波形上方标出了该乐音的起振、稳态和衰减阶段。划分这 3 个阶段的主要依据是频谱，因为在起振和衰减阶段频谱是不断变化

的，而中间部分的频谱相对稳定。但是，“稳态”并不意味着完全不变。例如，管风琴乐音在持续时间较长时，如图 5.1 所示的音符，其持续时间大约为 2s，音色听起来没有什么变化，但是，从图中的声压波形看出，即使在“稳态”阶段，幅度或包络仍然随时间变化。当合成乐音时，特别是当在取样合成器中使用循环技术时，认识到乐音的这一特征是非常重要的。

对管风琴的 principal 音管组而言，乐音的衰减阶段开始于手指松开琴键、来自风箱使“空气簧片”振动的气流停止之时。此图例中乐音的衰减大约持续 200ms，较高的谐波成分在第 1 次和第 2 次谐波之前就衰减掉了。如果录音是在封闭空间中进行的，那么要保持乐音的衰减特性就变得非常困难，因为封闭空间的特性正好与自由空间相反（见第 4 章），由房间产生的混响声会叠加到乐音的衰减特性上（见第 6 章）。在这个例子中要看到衰减阶段的详细信息比较困难，因为其持续时间较起振阶段长，很难将整个过程详细地在时间轴上描绘出来。

乐音的起振阶段对音色起非常重要的作用。同时，由于听音者在乐音的稳态阶段能较真切地感受其音色，因此，衰减阶段对音色感知所起的作用无疑比起振阶段和稳态阶段弱。起振阶段不易受演奏所在环境声学特性的影响，因为第一次反射声对直达声的染色效应（见第 6 章）往往发生在起振阶段结束以后（因此传递给听音者的是未经染色的声音起振信息）。可以肯定，人们在听到未染色的起振部分后反射声才到达。因此，起振阶段对音色感知是至关重要的，衰减阶段只是影响音色感觉的因素之一。当对乐音进行声学特性分析时，一个时间范围扩展到起振阶段的声谱图是极为有帮助的。

图 5.1 的下半部表示将时间轴放大后的起振阶段和稳态阶段的开始部分，其中起振阶段大约持续 70ms。图中可以清楚地看到各次谐波建立的过程，而且二次谐波比一次和三次谐波提前约 30ms 出现。这是管风琴带辅音起振音管的特点，在声学上表现为在起振阶段首先跳到第一或更高的超吹模式，开放式哨管的第一超吹模式是二次谐波（见第 4 章）。仔细聆听带有合成吹管特征的音管可以发现，开放式音管在开始时有瞬间的高八度音，因为其第一超吹模式是二次谐波；而封闭式音管则能听到瞬间的高一个八度加一个五度的音，因为其第一超吹模式是三次谐波。图

中四次谐波和三次谐波同时出现，在五次谐波出现时，四次谐波短暂地消失约 30ms，而七次谐波几乎与二次谐波同时出现，然后幅度消失约 60ms。由图可知，谐波成分的逐渐建立过程反映在声压波形上是其起振阶段包络的变化，尤其是幅度的逐渐增大。管风琴的 principal 音管组的起振阶段可以认为是由一系列声学事件或声学“线索（cues）”组成的，它们是目前人们所能得到的对乐音音色的感知起潜在作用的信息。

5.2.2 乐音的起振阶段

为了提供一些数据来说明乐音的起振阶段对音色感知的重要性，图 5.2 ~ 图 5.4 所示分别为 4 种管风琴音栓、4 种木管乐器和 4 种铜管乐器 C4（除了长号和大号为 C3）的起振阶段和稳态阶段开始部分的波形和声谱图。

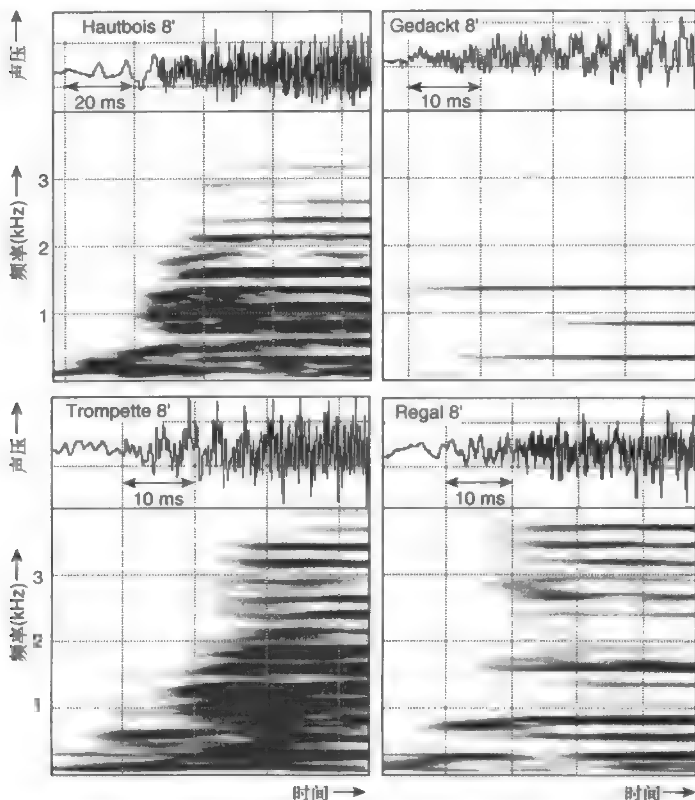


图 5.2 4 种管风琴音栓的 C4 起振阶段的波形（上）和声谱图（下）（hautbois 8'（簧管）、trompette 8'（簧管）、gedackt 8'（封闭式哨管）和 regal 8'（簧管））（hautbois 和 trompette 的 LTAS 见图 4.22，gedackt 的 LTAS 见图 4.17）

需要提醒的是，这些图形只是作为一些例子，说明这些乐器在起振阶段的一般特性。如果不同的演奏者用不同的乐器、在不同的地方演奏相同的音符，甚至是同一个演奏者用同样的乐器、在相同的地方演奏，同时保持所有细节不变，测得的波形和频谱都可能存在较大差异。

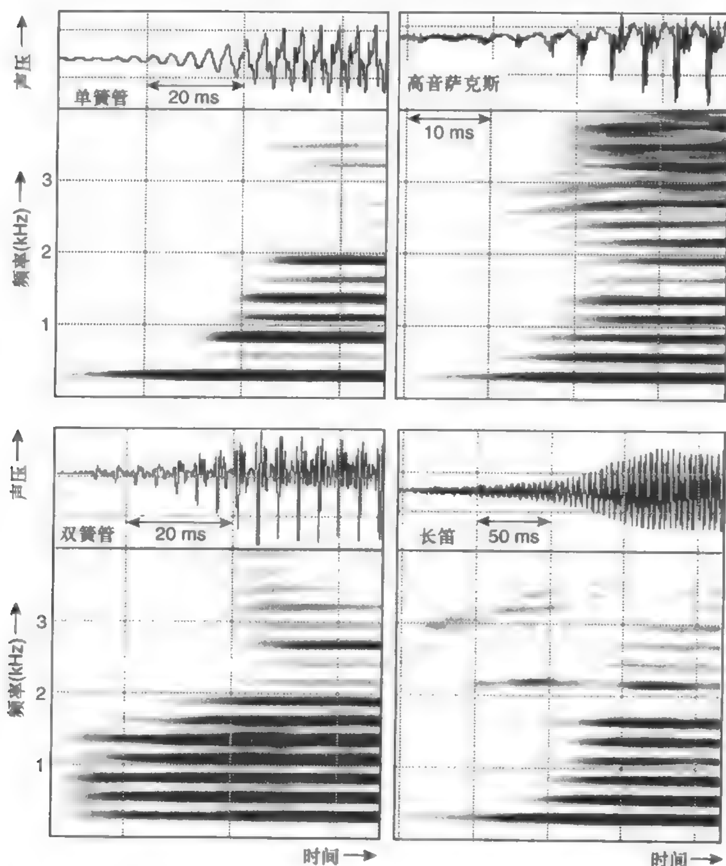


图 5.3　4 种木管乐器的 C4 起振阶段的波形（上）和声谱图（下）（单簧管、长笛、双簧管和高音萨克斯管）（单簧管和高音萨克斯管的 LTAS 见图 4.24）

图 5.2 所示的波形和频谱来自管风琴的 3 种簧管音栓 hautbois、trompette 和 regal 以及一种封闭式哨管音栓 gedackt，后者作为封闭式哨管音栓的例子。封闭式哨管仅支持奇次谐波（见第 4 章），从这个特定例子的起振阶段看出，五次谐波首先出现，它是第二超吹模式，其音高比基频高两个八度加一个大三度（参看图 3.3），随后是基频出现，然后是三次谐波，使该组音管具有独特的合成吹管特征。簧管组的起振阶段显

得复杂得多，因为每种情况都呈现更多的谐波成分。对 hautbois 和 regal 来说，基频首先出现；而对 trompette 而言，则是二次谐波首先出现。无论哪一种情况，在簧片振动的最初几个周期，基频呈上升趋势。对于特定的音管组，其各次谐波交错出现的不同状况是决定其音色的因素之一。trompette 谐波成分的频率上限几乎达到频率坐标的上限 4kHz，hautbois 的最高谐波频率可达到约 2.5kHz，regal 的四次和八次谐波基本上看不到（在本图设定的参考幅度情况下）。

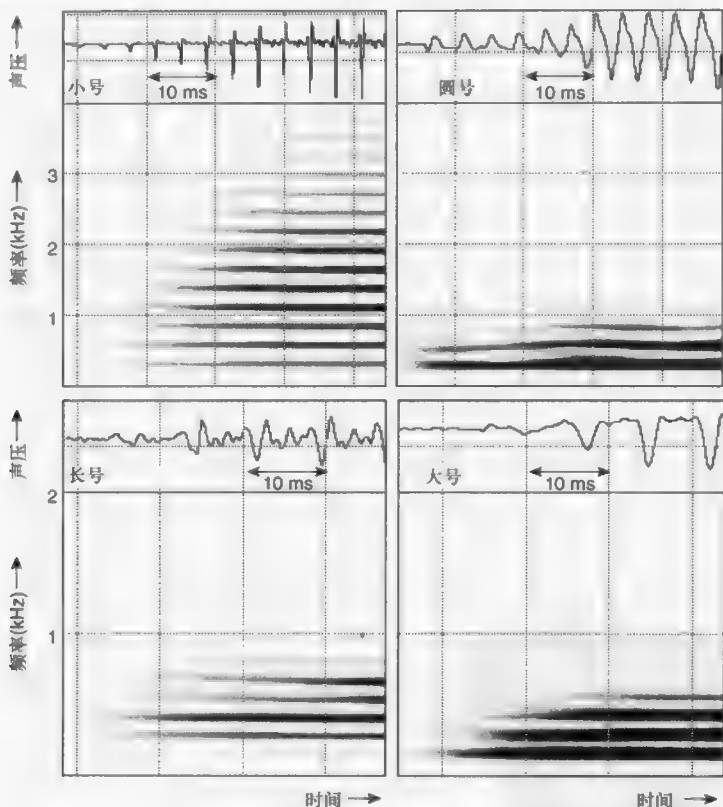


图 5.4　4 种铜管乐器的 C4（小号和圆号）和 C3（长号和大号）起振阶段的波形（上）和声谱图（下）（长号和大号的 LTAS 见图 4.29）

图 5.3 所示为 4 种木管乐器的波形和频谱，这 4 种木管乐器分别是单簧管、双簧管、高音萨克斯管和长笛。对单簧管和高音萨克斯管而言，基频首先出现；而双簧管则是二次谐波首先出现，大约 5ms 之后是三次和四次谐波，然后约 8ms 之后是基频。单簧管的高次谐波大约在基频之后 30ms 相继出现，其中奇次谐波的主导作用在第 4 章已经

讨论过了。对长笛这个特例来说，在基频之前出现的是明显的“呼吸声”。从声谱图上可以看到这个呼吸声频率在 2kHz 以上，持续时间大约是 70ms。在基频出现约 80ms 以后，高次谐波一个接着一个陆续出现，较长的起振阶段是长笛声音的特点。在起振阶段，波形的周期性逐渐形成，无论哪一种情况，都存在一个适当的时间间隔使振幅达到稳态时的振幅。

图 5.4 所示为铜管乐器的波形和频谱。小号和圆号演奏的乐音是 C4，长号和大号演奏的是 C3。小号是其中唯一的高次谐波能量较强的乐器，其四次、五次和六次谐波具有最高的幅度。图中其他乐器则不能明显地看到约四次以上的谐波成分（圆号和大号）或六次以上的谐波成分（长号）。所有 4 种乐器的起振阶段都是基频首先出现（注意长号看起来并不十分明显），随后是其他谐波成分，而且起振时间短，波形非常迅速地进入周期性阶段。

图 5.5 的上半部为小提琴 C4 的完整波形和声谱图。由图可知，颤音大约在乐音起始 250ms 后出现，主要表现为高次谐波频率的变化，这是因为使用了线性频率坐标。如果 f_0 变化 xHz，则意味着二次谐波频率变化为 2xHz，三次谐波频率变化为 3xHz 等，n 次谐波频率的变化就是 nxHz。因此，当图中频率坐标采用线性刻度时，颤音中高次谐波频率的变化要比低次谐波频率的变化明显得多。正如图中所示，颤音往往稍后出现，这和微妙的演奏技巧有关。图中所示小提琴乐音的起振阶段约为 160ms，衰减阶段约为 250ms。

在本节的最后，对来自 CD 录音的专业男高音（f_0=B$_b$4）的普契尼“Tosca”第 2 幕“Vittoria”的最后 3 个音节（即“toria”）进行分析（参看图 5.5 的下图）。这是乐队停止演奏、只留下男高音歌声的一段音乐，图中显示的是乐队停止 500ms 后的波形和频谱。声谱图左下角部分很难看到任何详细的声学形态，说明了在实际的音乐声学分析过程中会面对许多问题。声谱图清楚地显示出所包含的谐波成分以及颤音的频率波动程度，并且在频率为 2.4kHz ~ 3.5kHz 处可以明显地看到一个“歌唱者共振峰”（可与图 4.38 比较）。此演唱者的演唱显示了 3 个音节（“toria”）中的第 1 和第 3 音节时间较长，而第 2 音节时间较短，同时还可以看到，第 2 音节的所有谐波在声谱图的中部呈现一个谷点。

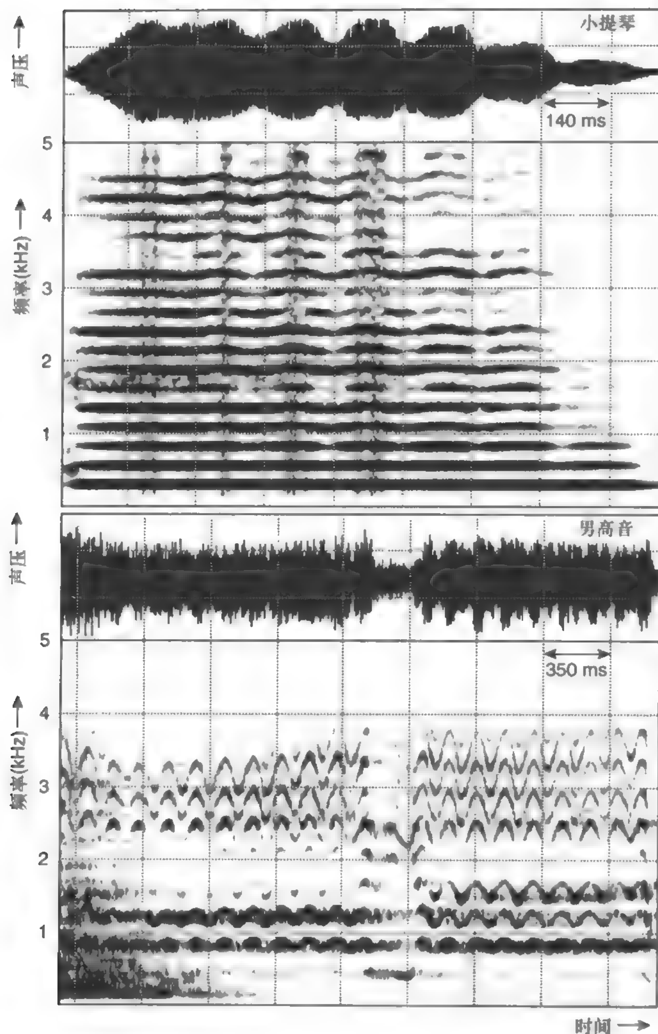


图 5.5　小提琴 C4（262Hz）的波形和声谱图（上）以及来自 CD 录音的专业男高音（f_0=B_b4）的普契尼"Tosca"第 2 幕"Vittoria"的最后 3 个音节的波形和声谱图（下）

5.3　音色的心理声学

为了了解听音者对不同乐器音色的感知以及决定音色的各种声学因素，研究人员已经进行了一些心理声学实验。从这些实验可以得出一些结论，其中之一是，当乐音的起振阶段和衰减阶段去除后，听音者很难

辨别不同乐器的音色。例如，如果把录制的小提琴的乐音和同一音符的小号录音的起振阶段和衰减阶段去除，则听音者很难将二者加以区别。图 5.1 ～图 5.5 给出了一些乐音起振阶段声学特性的例子，从中可以看出它们在起振阶段声学特性的差异。因此，在起始阶段弦乐器的弓弦摩擦声、铜管乐器辅音式的起振、吹笛者的呼吸噪声、簧片起振时的拍打声、钢琴琴锤发出的沉闷击打声以及拨弦键琴的琴锤最终回落在琴弦上的声音等都是辨别不同乐器声音的声学“线索”。当进行声学乐器声音合成时，为了获得听起来比较自然的音色，乐声的这些声学特征必须加以重视。

5.3.1　临界频带与音色

音色感知的心理声学基础是听觉临界带宽随频率变化的特性，它反映了听觉的频谱分析特点。临界带宽随频率的增大而增大，关于其变化规律的一般性结论见第 3 章中音高感知一节（参看 3.2 节）。因此，无论乐音的基频 f_0 是多少，其第 5 ～ 7 次以上的谐波成分一般不能被听觉分辨出来，而第 5 ～ 7 次以下的谐波成分则能够被听觉一一分辨出来（参看图 3.11），说明这些谐波成分各自都对音色感知起着显著作用。另一方面，由于听觉不能分辨第 5 ～ 7 次以上的谐波成分，因此它们各自对音色感知不可能有多大影响，但是，作为处在某个临界频带内的一组谐波成分，则可能对音色感知产生一定的影响。在了解了这些性质的基础上，下面将对那些声学特性在本书做了介绍的乐器音色重新进行分析。但是要记住，这些分析结果是针对特定的演奏者在特定的声学环境和特定的乐器上演奏某个特定响度和音高音符的结果，任何一点改变都可能对结果产生影响。

在已经展示频谱图的乐器中，其稳态频谱包含显著的五次到七次以上谐波成分的乐器有管风琴簧管组（见图 4.22 和图 5.2）、高音萨克斯管（见图 4.24 和图 5.3）、小号（见图 5.4）、小提琴和专业歌声（见图 5.5）。这些乐器的音色可以用“明亮的”、“辉煌的”或者“刺耳的”等音质评价术语描述，以区别其他乐器的音色。稳态频谱不包含显著的五次到七次以上谐波能量的乐器有管风琴的 principal 8' 音管组（见图 4.17 和图 5.1），gedackt 8' 音管组（见图 4.17 和图 5.2），单簧管、双簧管和长笛（见图 4.24

和图 5.3）以及长号、圆号和大号（见图 4.29 和图 5.4）。和其他管风琴音管组或其他同类乐器（木管或铜管）相比，它们的音色可以用“不明亮的”或“暗淡的”、“不辉煌的”或“单调的”和“不刺耳的”或“柔和的”等术语描述。

在上述的后一组乐器中，还可以针对音色做进一步的分类。一类是所包含的谐波高达五次到七次谐波，如单簧管、双簧管和长笛；另一类是仅包含少量低次谐波，如管风琴的 principal 8' 音管组、gedackt 8' 音管组、长号、圆号和大号。长笛和单簧管、双簧管分在同一组可能很奇怪，但是，和单簧管、双簧管相比，长笛的七次谐波缺失（见图 5.3）是关键所在。七次谐波是听觉不能分辨的最低谐波，在六次或八次谐波存在的情况下，没有七次谐波的声音听起来比有七次谐波的声音柔和得多。而单簧管虽然存在七次谐波，但其六次和八次谐波都很弱，因此听起来也不那么刺耳。由于单簧管的奇次谐波起主导作用，所以单簧管具有其独特音色，常被描述为具有“鼻音性的”。对 tierce($1\frac{3}{5}'$) 和罕见的 septième($1\frac{1}{7}'$) 音管组声音熟悉的管风琴演奏家会充分认识到五次和七次谐波分别对音色所起的重要作用以及当与其他音管组同时使用时往往产生尖利刺耳的声音效果。

通过击打棒、膜或板发声的打击乐器（见 4.4 节）具有独特的音色，这主要是因为其自然振动模式频率之间的非谐波关系，为这一乐器家族音色的独特性提供了明确的声学“暗示”，使这类乐器听起来具有“铿锵”或“叮当”声，赋予了这类乐器独有的音色特点。

5.3.2　声学“线索”与音色感知

音色评判具有很强的主观性，因此评判结果存在个人差异。在对音高或响度进行评判时，听音者能够分别将声音从高到低或从弱到强进行排列。但在音色评判时，没有所谓的正确答案，听音者往往被要求聆听比较不同声音的音色，并在两个相反的形容词之间进行打分，例如“明亮的”（1）到“暗淡的”（10）、“华彩的”（1）到“单调的”（10）。一般需要几对这样的形容词对特定声音的音色进行评价。一定数量听音者评价结果的平均值代表对某一声音音色的总体评价结果。Hall(1991) 指出，

在一天之内完成的不超过 5 对评价项目的评价结果，从理论上说“足以精确地分辨任何不同的音色”。

研究人员试图了解乐器输出的某些声学特征与音色之间的关系。Grey（1977）曾经做过一个重要的实验。听音者被要求对两个合成乐器声音的相似度进行评价，并用 1 ~ 30 之间的一个数值度量。所有的声音都经过音高、响度和持续时间的校准，评价结果在一个多维坐标系中表示出来。这是一种根据听音者的相似度评价结果，将各种乐器间的相互关系用多维坐标空间中的相对位置来表示的一种计算技术。在 Grey 的实验中采用的是三维坐标空间，针对沿着坐标轴排列的各乐器的不同输出声学特性，对各坐标轴上产生主观评价距离的各种声学要素进行分析。Grey 对三个坐标分别赋予下述声学含义：（1）在“能量谱分布”图上高频成分越来越强；（2）“起振和衰减阶段高次谐波出现和消失的同步性”，从起振阶段所有谐波成分几乎同时出现到相继缓慢出现；（3）从带有“起振阶段优先出现的高频低幅度非谐波成分”的声音到那些在起振阶段不带有高频谐波能量的声音。这些研究表明：（a）关于音色人们能够并且已经进行了一些有用的实验工作；（b）可以得到一些与其他观察到的结果相符合的声学结论，例如强调与乐音起振阶段有关的 Grey 坐标（2）和坐标（3）对音色所起的作用。

乐声总是在不断变化着，即使在容易引起误解的所谓“稳态”阶段也是这样。这一点可以从关于小提琴和歌声的波形和声谱图的图 5.5 中清楚看出。由于管风琴气流相对比较稳定而有规律，因此管风琴乐音通常被认为是稳定的。然而从图 5.1 看到，即使在一个音符的稳态阶段，管风琴的乐音也呈现一定的振幅波动性。当进行乐音电子合成时，如果不在乐音的稳态阶段把幅度的波动反映出来，则在音色上会产生明显的差异。这从听感上清楚地说明了这种波动性的真实存在。为了吸引听者的注意力，任何声音都需要伴随着某种变化。当通过声音传递新的信息给听者时，无论是语声、音乐、环境声，还是自然的或人为的报警声，都需要声音信号以某种方式随时间变化，产生变化的可以是声音的音高、响度或音色。一些正在流行的声音后期制作效果，特别是合唱效果（见第 7 章），无论它们是以声音合成器的功能加入，还是以演播室效果器的形式加入，都充分说明了这一点。然而，虽然通过使声音随时间变化的后期处理可

以使声音听起来更有吸引力，但是，这种处理一般对改善合成声的整体自然度没有好处。

任何乐器乐音的音高、响度和音色在声音持续的整个过程中都处于动态变化中。音高和响度的主观度量是从“低”到“高”一维的，并且与某些能够实际测量到的物理量之间有相当直接的关系。但是，音色却没有这种一维的主观度量。有人提出用跟踪各谐波成分变化规律的方法来描述音色的动态变化特性，Pollard 和 Jansson(1982) 提出的“三色图”(“tristimulus diagram”) (译者注：这里借用色彩学中调色板的概念描绘音色) 就是方法之一，如图 5.6 所示。图中将每个乐音的时间过程在三角形图中描绘出来，其描绘方法是基于以下各谐波成分的能量：(1) 二次、三次和四次谐波能量，在图中用“中频”表示 (y 轴)；(2) 五次和五次以上的谐波能量，图中用“高频”表示 (x 轴)；(3) 基频 f_0(x 和 y 均为零处)。图 5.6 的各角落分别标有“中频”、“高频”和“f_0”，用来代表上述含义。因此，“三色图”中的某一点表明了 f_0、能够被听觉分辨的谐波能量和不能被听觉分辨的谐波能量三者之间的关系。

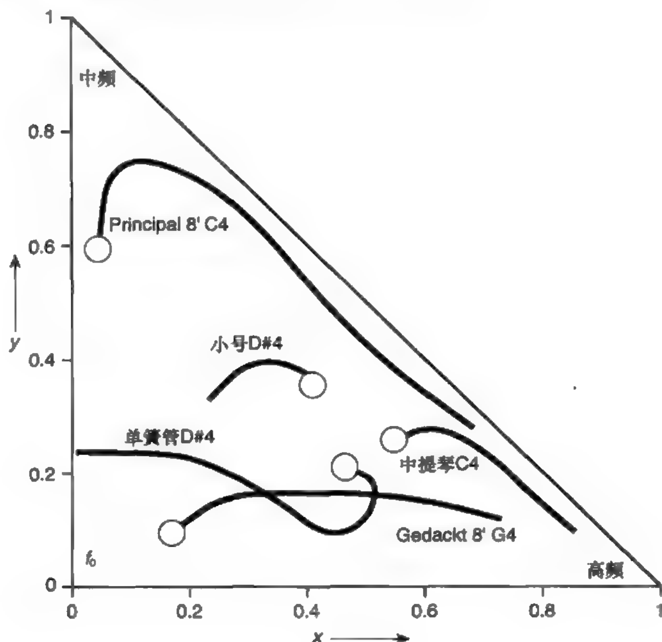


图 5.6　一些乐器乐音起振过程的“三色图”的音色表示 (每个乐音沿着线条轨迹到达用空心圆表示的稳态。“中频”代表“强的中频分量”；“高频”代表“强的高频分量”；“f_0”代表“强的基频”。数据来自 Pollard 和 Jansson 1982)

"三色图"使"中频"、"高频"和"f_0"之间的动态关系可以用一条轨迹直观地表示出来，一些乐器的乐音起振过程如图 5.6 所示（Pollard 和 Jansson 1982）。可见，时间过程并不平直。为了能够看得更清楚，没有对各条曲线进行校准。每个乐音的稳态用空心圆表示，曲线的另一端代表乐音的开始。这些乐音的持续时间如下：gedackt（10 ~ 60ms）、小号（10 ~ 100ms）、单簧管（30 ~ 160ms）、principal（10 ~ 150ms）和中提琴（10 ~ 65ms）。图中每个乐音的轨迹有很大不同，稳态落在不同的位置。Pollard 和 Jansson 对更多的上述乐器的乐音进行了分析，表明不同的乐器大致上对应于"三色图"中的特定位置。这为各乐器的音色差异提供了一个看得见的、基于临界频带分析的表示方法，它也提供了表示乐音形态特征的独特方法，可用于表示不同乐音音色上的差异。

"三色图"基本上给出的是乐音的 f_0 分量、f_0 以外的能够被听觉分辨的谐波分量和不能被听觉分辨的谐波分量三者之间的权重关系。关于音色的"三色图"表示法背后的想法并不是全新的。Helmholtz 曾在其 1877 年首次出版的、对学科发展有重要影响的学术著作《乐音的感知》（英文版于 1954 年出版）中提出关于音色的 4 条一般规律，"用来表示音色和乐音的各频率成分组成模式之间的关系"（Helmholtz 1877）。然而，当时并没有临界频带、如今的音高和响度感觉模型的相关理论去支持 Helmholtz 的想法。尽管如此，Helmholtz 关于音色的 4 条一般规律对于当今理解音色的客观性质同样具有重要意义。这 4 条一般规律阐述如下（参考本书 CD 第 63 条）。

（1）简单的乐音，例如音叉在共鸣室产生的声音和粗大的封闭式管风琴音管产生的声音。一般听起来比较柔和、令人感到愉悦，不存在任何粗糙感，但是强度欠缺，并且低音显得单调。

（2）包含有中等强度的约 6 次以下的谐波分量的声音听起来更和谐、更具有乐感，并且与简单乐音相比，音色丰富而华丽；在缺失高次谐波分量的情况下，其音色十分甜美、柔和。产生这种声音的乐器有钢琴、开放式管风琴音管、人声中较轻柔的钢琴般的声音和圆号。

（3）如果声音只包含奇次谐波分量（如狭窄的封闭式管风琴音管、在中部击打钢琴琴弦、单簧管等发出的声音），则音质比较空洞；当包含

大量的奇次谐波分量时，声音鼻音较重；当基频较强时，音色比较丰富；当基频与高次谐波相比不够强时，音质则较差。

（4）当高于六次或七次的谐波很显著时，音色比较刺耳和粗糙。……属于这类声音的乐器有弓弦乐器、有簧音管、双簧管、巴松、簧风琴和人声，粗糙、高亢的铜管乐器的声音极具穿透力，因此，与其他类似音色的较柔和的乐音相比，更适合于产生力量感。(Helmholtz 1877 的英文版，第 118 ~ 119 页，1954 年。)

这些一般规律为根据声音频谱判断音色提供了基本依据，而频谱又与当今人们对临界频带的理解直接相关。第 2 条和第 4 条基本规律对不包含高于六次或七次谐波分量的声音和包含高于 6 次或 7 次谐波分量的声音进行明确区分，第一条则强调基频 f_0 的重要性。值得注意的是，这些结论与“三色图”的坐标轴以及 Grey 的多维坐标实验中的坐标轴 I 的结论很相似。还要注意到，Helmholtz 选用了哪些术语描述不同声音的音色（请记住 Alexander Ellis 已在 1954 年将其从德文翻译成英文）。作为 Timbre(音色)的另一种表达，Alexander Ellis 将德语中的“klangfarbe”译为 quality of tone(音质)。因为他认为这个词“在我们的语言中更通俗易懂”。他注意到也可以选用“register(音域)”、“colour(音色)”或“timber(音色)”中的某一个，但是他没有这样做，因为当时这 3 个词还有其他含义。

（1）Register——“在声乐中已有明确的含义，不应该去混淆它。”

（2）Timbre——有定音鼓、头盔、带头盔的臂护等含义，在法语中有邮票的意思。因此，Ellis 认为：“Timbre 是个外来语，经常出现发音错误，因此不适合在这里使用”。

（3）Colour——“常在音乐中作为一种比喻表达”。

Howard 和 Tyrrell(1997)提出了一种基于声谱图的 4 条一般规律的频域表示法，而这种频谱图反映了听觉模型信号分析系统的输出（译者注：模拟听觉的频谱分析特性，在下文中称为听觉模型声谱图），这部分内容将在下面几个段落介绍。图 5.7 所示为下述 4 种合成声的听觉模型声谱图：(a)正弦波；(b)只包含前 5 次谐波的复音周期信号；(c)只包含第 1 ~ 19 次奇次谐波的复音周期信号；(d)包含前 20 次谐波（偶次和奇次）的复音周期信号。在每种情况下，f_0 在 128Hz ~ 160Hz 之间变化。声音 a ~ d 分别代表 Helmholtz 在一般规律 1 和 4 中所述的声音。关于听觉模型声谱

图需要注意的关键问题是，由于声音是通过一系列听觉滤波器进行处理的，听觉滤波器具有特定的滤波器形状和带宽，因此，频率轴是根据 ERB（equivalent rectangular bandwith，等效矩形带宽）进行刻度的，即采用对数频率刻度，使各频带在声谱图的频率轴上是等间隔的。

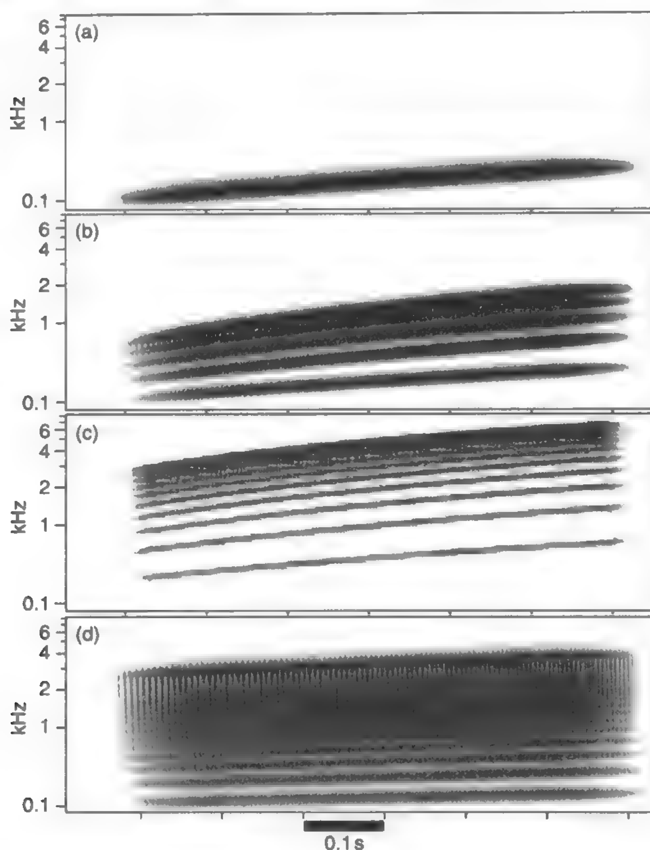


图 5.7 f_0 在 128Hz ~ 160Hz 之间变化的合成声的听觉声谱图：（a）正弦波；（b）只包含前 5 次谐波的复音周期信号；（c）只包含第 1 ~ 19 次奇次谐波的复音周期信号；（d）包含前 20 次谐波（偶次和奇次）的复音周期信号

Howard 和 Tyrrell（1997）指出，正弦波的听觉模型声谱图表现为水平方向的一条能量带（f_0 分量），其中频率随时间逐渐升高（从 128Hz ~ 160Hz）。请注意在上述 4 个声谱图中都存在基频，并且基频的变化是相同的。对于包含前 5 个谐波的声音而言，图中可以清晰看到 5 条水平能量带，并且由于听觉滤波器等效带宽 ERB 的特性，使得这些能量带随着频率升高越来越靠近。在第 4 种声音的听觉声谱图中，由于包含所有高达 20 次的谐波分量，最低的 6 或 7 个谐波可以被听觉分辨出来，表现为

水平能量带，但是，高于 7 次的谐波能量表现为垂直线条，称为“沟痕”，大约每一个周期出现一次（请回忆前述的音高感知时间论）。这是因为所有中心频率高于 7 次谐波的听觉滤波器的带宽都大于声音的 f_0，因此，每个听觉滤波器都至少通过两个相邻的谐波，由此产生拍音现象。在第 3 种声音的声谱图中，由于仅包含高达 19 次的奇次谐波成分，可以看到大约 7 条可分辨的水平能量带，分别对应于最低的 7 个奇次谐波，即第 1 ~ 13 次奇次谐波。听觉不能分辨的频率临界点对应于 ERB 带宽大于相邻谐波间隔的频率（两个相邻奇次谐波的间隔为 $2f_0$）。对于第 3 种声音而言，这一点大约就在声音的第 5 ~ 7 个谐波频率之间，即在第 10 次与第 14 次谐波之间的某一点。

表 5.1 的数据来自 Howard 和 Tyrrell(1997)，分别针对 Helmholtz 的 4 条一般规律列出了听觉声谱图的主要声学特征、若干音色评价术语（包括那些 Helmholtz 提出的）和一些乐器例子。然而，Howard 和 Tyrrell 指出，"乐器只是大致上进行分类。例如，当采用扩展的演奏技巧进行演奏时，根据其频域（听觉模型）的特性，一些乐器的声音很可能会落入另一个音色组。" 听觉模型声谱图分析法也已经应用于语声（Brookes etal，2000）、与法庭有关的音频信号处理（Howard et at，1995）和歌声（Howard 2005），它为探究与听音感觉相关的声音的声学特性提供了新的视角。

表 5.1　针对 Helmholtz（1877）的 4 条一般规律列出的代表声音频域特性的听觉模型声谱图、若干音色评价术语和一些乐器例子（Howard 和 Tyrrell 1997）

Helmholtz 规律	听觉模型频谱图	音色术语	乐器
1	基频 f_0 为主	纯净的	音叉
		柔和的	粗大的封闭式管风
		琴哨管	
		简单的	
		愉悦的	巴洛克长笛
		低音单调的	
		无粗糙感的	
2	低次谐波为主	甜美和柔和的	

续表

Helmholtz 规律	听觉模型频谱图	音色术语	乐器
		丰富的	圆号
		华丽的	现代长笛
		暗淡的	竖笛
		管	开放式管风琴哨
		单调的	
		不太刺耳的	
		平淡的	轻柔的歌声
3	奇次谐波为主	空洞的	单簧管
		鼻音的	狭窄的封闭式管风
		琴哨管	
4	沟痕为主	尖刺的	单簧管、巴松
		粗糙的	小号、长号
		明亮的	管风琴哨管
		光辉的	高声歌唱
		高声调的	弓弦乐器
		花哨的	管风琴簧管

为了更全面地了解音色，还有许多工作要做。虽然已得到一些关于乐器的哪些声学特性对音色差异起决定性作用的研究结果，但是，这些结论还比较粗浅，难以解释一个训练有素的听音者如何能够判别两个声音的音色差。例如，由两个不同的制作者制造的两把小提琴的音色差。从被称为乐器之王的管风琴可以看到，即使在人们对听觉系统缺乏详细了解的过去，人们对音色在音乐演出中重要性的认识也已经有几百年的历史了。

5.4 管风琴作为音色合成器

管风琴是最早形式的用于产生不同音色声音的“添加式合成器”之一。添加式合成器通过将各种谐波成分加在一起来控制输出声的音色，管风琴的音栓或拉栓就是用来进行这样的处理。有文献指出，最早的管风琴

出现在至少公元前 250 年（Sumner 1975）。

现代管风琴一般由两个或更多的键盘以及一个踏板盘组成。每个键盘音域可高达 5 个八度（从 C2 ~ C7 共 61 个琴键），踏板盘的音域可高达两个半八度（从 C1 ~ G3 共 32 个键）。大型管风琴可能有多达 6 个键盘，小型管风琴可能只有一个键盘，同时没有踏板盘。5 个八度键盘上的每个音栓控制 61 根音管，另外还有一种称为“混合（mixtures）”音栓，每个琴键控制几根音管。一组具有和钢琴相同基频（即 A4 的基频为 440Hz，参看图 3.21）的音管通常在键盘上用“8 英尺”（1 英尺 =0.304 8m）或 8' 表示，在踏板上则用“16 英尺”或 16' 表示。8 英尺和 16 英尺分别对应于键盘上的最低音（C2）和踏板的最低音（C1）的开放式音管的长度。因此，踏板的声音比乐谱上的音符低一个八度。4' 音管组和 2' 音管组分别比 8' 音管组高一个和两个八度，而 32' 音管组比 16' 音管组低一个八度。值得注意的是，这种以英尺度量的长度术语只用来表示该音管组的音调，而并不说明音管是开放式还是封闭式。因此，一组和 8' 开放式音管相同音高的封闭式音管的最低音音管的长度是 4ft，但却用 8' 表示。

管风琴的每个键盘都设有不同尺数（footages）的若干个音栓，大多数是由哨管组成的。有些音栓专门用于独奏，音高通常设在 8' 上；但大多数音栓用于混合演奏，通过拉出不同的音栓组合，使声音和谐地混合，从而产生不同的响度和音色。音色的控制是通过加强键盘上的 8'（踏板上 16'）自然谐波系列达到的。下列公式给出音栓的尺数与其基频所要加强的谐波次数之间的关系：

$$\text{音栓尺数} = \text{第一次谐波的尺数}/N \qquad (5.1)$$

其中，N 为所要加强的谐波次数（1，2，3，4…）。

例如，基频加强第 6 次谐波（属于 8' 谐波系列）的音栓尺数为：

$$\left[\frac{8}{6}\right] = 1\frac{2'}{6} = 1\frac{1'}{3}$$

基频加强踏板上的第 3 次谐波（与 16' 谐波系列相关）的音栓尺数为：

$$\left[\frac{16}{3}\right] = 5\frac{1'}{3}$$

例 5.1 试计算加强 8' 谐波系列的第 3 次和第 7 次自然谐波的音栓尺数。

加强第 3 次谐波的音栓尺数为：$\left[\frac{8}{3}\right]=2\frac{2}{3}'$

加强第 7 次谐波的音栓尺数为：$\left[\frac{8}{7}\right]=1\frac{1}{7}'$

表 5.2 列出了管风琴音栓上可能给出的尺数，同时给出了一部分长笛家族成员、其他哨管音栓和簧管音栓的名称。与 8' 音栓的音程差用大音阶的阶数表示，不同尺数的音栓的基频所加强的 8' 和 16' 谐波系列用谐波次数表示。那些所加强的谐波既不是基频 f_0，也不是 f_0 的整数倍八度关系（即 2 ：1，4 ：1，8 ：1…2^n ：1）的音栓称为“变化（mutation）”音栓。表中所列的音栓既有键盘上的音栓，也有踏板上的音栓。只有大教堂里的大型管风琴才包含上述大多数音栓，而许多牧区小教堂的管风琴只包含其中一小部分音栓，通常只包含 8'、4'、2' 哨管音栓和一个或两个簧管音栓，有的可能还包含 $2\frac{2}{3}'$ 的哨管音栓。

管风琴演奏者通过组合设置在键盘上的加强 8' 谐波系列和踏板 16' 谐波系列的音栓可以塑造出管风琴的总体声音效果（参考本书 CD 第 64、65 条）。然而必须注意到，由开放式哨管组成的 8'diapason 或 principal 音栓不仅具有相同的 8' 基频，而且本身就包含丰富的谐波成分（参看图 5.1）。加入一个开放式哨管 4'principal 音栓后，不仅加强 8' 音栓的二次谐波，而且加强其他所有偶次谐波。一般来说，当增加某个音栓加强某次（n 次）谐波时，它同时也加强第 $2n$、$3n$、$4n$……次谐波。

当采用十二平均律调音系统（见第 3 章）时，管风琴音色上存在一个问题。由于变化音栓要以整数比的频率关系加强自然谐波成分，因此就不能采用十二平均律调音，否则当演奏和声时，就会产生拍音。例如，当 $2\frac{2}{3}'$ 音栓和 8' 音栓同时演奏 C 时，$2\frac{2}{3}'$ 音栓 C 音的 f_0 与 8' 音栓 C 音的三次谐波频率相同；但是，如果此时同时演奏 G 音形成两个音符的和声，则 8' 音栓 G 音的基频就会与 $2\frac{2}{3}'$ 音栓 C 音的 f_0（和 8' 音栓 C 音的三次谐波）产生拍音。因此，如果管风琴采用十二平均律调音，则会对加入“变化”音栓的管风琴声音产生拍音的染色效果。这就是随着平均律调音系统的出现，变化音栓越来越少使用的原因（Padgham 1986）。

表 5.2　一些非混合音栓的音栓名称、尺数、踏板或键盘的最低 C 的 f_0、用大音阶的阶数表示的与 8' 的音程和基频所加强的 8' 和 16' 谐波系列的谐波次数（粗体表示哨管音栓，斜体表示簧管音栓）

音栓名称 （长笛类、其他哨管、簧管）	尺数	与 8' 的音程	8' 谐波次数	16' 谐波次数	最低 C 的 f_0
Gravissima、resultant bass *Duble trombone*	64'	之下第 22 阶	–	–	8.175Hz（踏板）
Double bourdon、acoustic bass double open diapason、 Contra violone contra *Trombone、contra bombarde*	32'	之下第 15 阶	–	–	16.35Hz（踏板）
Bourdon、open bass、open wood Quintadena、grosse geigen、 double diapason、 *regal、posaune、ophicleide、* *bombarde、fagotto、trombone、* *double trumpet*	16'	之下第 8 阶	–	1	130.8Hz
Double quint	$10\frac{2}{3}$'	之下第 5 阶	–	–	196.0Hz
Stopped diapason、gedackt、 claribel、open diapason、principal、 dulciana、gamba、salicional、 *tubatrumpet、tromba、horn、* *oboe、cornopean、krummhorn、* *clarinet、posaune*	8'	相同	1	2	261.6Hz
Voix celeste、vox angelica	8'–#	相同	1	2	270.0Hz
Quint	$5\frac{1}{3}$'	之上第 5 阶	–	3	391.0Hz
Rohrfl ü te、spitzfl ü te、 principal、gemshorn、octave、salicet *clarion、octave tuba、schalmei*	4'	之上第 8 阶	2	4	523.3Hz
Double tierce、tenth	$3\frac{1}{5}$'	之上第 10 阶	–	5	659.3Hz
Rohrquint、nazard、twelfth octave quint	$2\frac{2}{3}$'	之上第 12 阶	3	6	784.0Hz
Piccolo、principal、fifteenth、 super octave	2'	之上第 15 阶	4	8	1 047.0Hz
Tierce、seventeenth	$1\frac{3}{5}$'	之上第 17 阶	5	10	1 319.0Hz
Larigot、nineteenth、nasat	$1\frac{1}{3}$'	之上第 19 阶	6	12	1 568.0Hz
Septi è me	$1\frac{1}{7}$'	之上第 21 阶	7	14	1 831.8Hz
Sifflöte、spitzflöte	1'	之上第 22 阶	8	16	2 093.5Hz
None	$\frac{8}{9}$'	之上第 23 阶	9	18	2 349.3Hz

如今在回归纯正的早期音乐浪潮的推动下，管风琴音乐重新受到欢迎。在管风琴中趋向于不使用平均律调音系统，这样就能够使用更多的"变化"音栓，这尤其赋予复调音乐以新的生命力。

一般大型管风琴中都设有比 8'（键盘）和 16'（踏板）的基频低一个八度的音管组。因此，在键盘上可能会找到 16'bourdon、double open diapason 或 double trumpet 音栓，在踏板上可能看到 32'contra open bass 或 double open diapason 音栓。除了美国费城的约翰·沃纳梅克管风琴和悉尼百年纪念大厅的管风琴外，很少会看到比 8'（键盘）和 16'（踏板）低两个八度的音栓，例如，键盘上的 32'contra violone、踏板上的 64'gravissima、contra trombone。

加强高次谐波的音栓在管风琴中更为罕见，其中 septième 和 none 是最为罕见的两种，通常用于增加音色的"尖刺感"。tierce 产生的声音极其尖利刺耳，有时在法式管风琴中能够看到。为了便于对所塑造声音的响度和音色进行控制，一些加强高次谐波的音栓组合在一起，形成所谓的"混合"音栓。这些混合音栓的每个琴键至少控制两组音管，分别加强不同的高次自然谐波。表 5.3 列出了一些常用的混合音栓。

表 5.3　常见的混合音栓名称、音管组数目和用大音阶的阶数表示的与自然谐波系列（键盘的 8' 和踏板的 16'）第一次谐波的典型音程（注意正如文中提到，音程会随键盘或踏板上的不同位置而变化）

混合音栓名称	音管组数目	与第一次谐波的典型音程（阶数）
Sesquialtera	II	12、17
Tertian	II	17、19
Mixture	II	19、22
Zimbel	III	15、17、19
Mixture	III	19、22、26
Cymbal	III	29、33、36
Scharff	III	26、29、33
Plein jeu/mixture	IV	19、22、26、29
Mixture	V	15、19、22、26、29
Pedal mixture	V	12、17、19、21（降半音）、22
Kornet/cornet	V	1、8、12、15、17

在声音合成过程中，混合音栓还起着另一个重要作用。当仅仅在混合音栓上演奏半音音阶时，所听到的半音音阶不会保持连续。在每隔 1

个八度或第 12 个半音会听到音高的跳跃，这是因为混合音栓的音域比键盘的 5 个八度音域窄得多。因此，混合音栓只能对键盘低音区的声音增加明亮色彩，而对高音部分的效果则基本上听不到（音管长度太短难以实际应用）。由 4 组音管组成的混合音栓可以是从 C1 ~ C2 的 19、22、26、29， 从 C2 ~ C3 的 15、19、22、26， 从 C3 ~ C4 的 12、15、19、22，从 C4 ~ C5 的 8、12、15、19 和从 C5 ~ C6 的 8、12、15、15。表中的数值表示用大音阶的阶数表示的最低 C 的基频与第 1 次自然谐波的音程。由于包含音程差为 21（降半音）的音管组（对应于第 7 次谐波），表中的“pedal mixture”音栓会产生很强的簧管音色。

由于 Sesquialtera 和 Tertian 都包含 Tierce（第 5 次谐波），因此常用于合成独奏音栓。Cornet 也常用作独奏音栓，特别是用于教堂礼拜仪式前后或进行之中的演奏。如果没有这种音栓，它还可以由以下音栓合成：8'、4'、$2\frac{2}{3}'$、2' 和 $1\frac{3}{5}'$，或者 8'、4'、2' 和 Sesquialtera（如果有这些音栓）。

5.5 听觉的“欺骗”效应

本节将介绍一些从某种意义上说会对听觉产生“欺骗”的声音。这些声音的共同特点是，它们在听觉上会产生某种听感，而又找不到与之相对应的物理量或声学量。换句话说，声音主观感觉和客观存在并不总是能直接对应起来的。虽然所给出的一些例子可能对演奏者或作曲家没有明显的帮助，但是，它们将来有可能在电子音乐制作中得到应用，尤其当能够精确控制输出声音的各种声学参数时，可以用来产生某种特殊的音乐效果。

5.5.1 纯音的听感

当同时聆听两个纯音时，听觉并不是总能分别听到这两个纯音。第 2 章中针对图 2.6 关于临界频带的阐述提供了声音在某种意义上“欺骗”听觉的一个例子。只有当两个纯音的频率差大于临界频带时，听觉才能分别感觉到它们的存在。否则，听觉听到的是一个粗糙的融合音或者一个

拍音，这取决于这两个纯音的频率差。

当同时倾听两个纯音时，可能还会听到比这两个纯音频率都低的其他频率的声音，这些较低频率的声音在声学上并不存在于激励声中。这种同时倾听至少两个纯音时在听觉上产生的其他声音，称为“结合音”。最容易听到的结合音是两个纯音的频率差（较高频率减去较低频率），称为“差频”：

$$f_d=f_h-f_l \tag{5.2}$$

其中，f_d 为差频频率，f_h 为较高纯音的频率，f_l 为较低纯音的频率。

请注意，当频率差小于约 12.5Hz 时，就会产生拍音（见第 2 章）。其他由两个纯音产生的结合音频率可按如下公式计算：

$$f_{(n)}=f_l-[n(f_h-f_l)]=f_l-nf_d \tag{5.3}$$

其中，$f_{(n)}$ 为第 n 次结合音频率，$n=1，2，3，4，\cdots$，f_l 为较低纯音的频率，f_h 为较高纯音的频率。

这些结合音的频率总是低于较低纯音频率，差值为差频的整数倍。没有人能听到所有的结合音，有些人一个结合音都听不到。最容易被听觉感知的是差频、第一次和第二次结合音（n=1 和 n=2），也称为“第二次差频”和“第三次差频”（Rasch 和 Plomp 1982）。

例 5.2 试计算当频率为 1 200Hz 和 1 100Hz 的两个纯音同时倾听时产生的差频和前 4 个结合音频率。

由式（5.2）得，差频频率为：f_h-f_l=1 200−1 100=100Hz

由式（5.3）得，前 4 个结合音频率为：

当 n=1 时，$f_{(1)}$=1 100−1×100=1 000Hz

当 n=2 时，$f_{(2)}$=1 100−2×100=900Hz

当 n=3 时，$f_{(3)}$=1 100−3×100=800Hz

当 n=4 时，$f_{(4)}$=1 100−4×100=700Hz

当两个纯音为同一基频 f_0 的相邻两个谐波（在例 5.2 中，它们分别是 100Hz 的第 11 次和第 12 次谐波）时，差频为 f_0，其结合音形成该系列谐波缺失的谐波成分；而当两个纯音不是某谐波系列的成员时，所产生的结合音与它们之间没有相同的基频 f_0，但是，结合音之间的频率间隔是相等的。

当两种乐器产生的声音非常接近纯音，基频较高并且靠近时，例如高音竖笛、巴洛克长笛或短笛，产生的结合音很容易被听觉感知。

当两个纯音正好是以其差频为基频的谐波系列的两个相邻谐波时，结合音是与较低音相邻的、在较低音之下的、连续的谐波成分（即差频为基频 f_0，结合音为差频的整数倍）。因此，结合音与两个音符的音乐关系取决于它们所采用的调律系统。当采用平均律调音系统时，两个音的音程肯定不是完全音，因此其结合音只是非常接近于、但决不正好是一系列谐波成分。

例 5.3　如果两只和音竖笛同时演奏平均律的 A5 和 B5，请问平均律上的哪个音接近于最容易感知的结合音？

最容易感知的结合音是差频和式（5.3）中 $n=1$ 和 $n=2$ 时的两个结合音。由图 3.21 可查出这两个音符的基频 f_0 分别是 880.0Hz（A5）和 987.8Hz（B5）。因此：

差频为：987.8-880.0=107.8Hz，最接近的音符是 A2（f_0=110.0Hz）。

结合音为：

当 $n=1$ 时，880.0-107.8=772.2Hz，最接近的音符是 G5（f_0=784.0Hz）；

当 $n=2$ 时，880.0-2×107.8=664.4Hz，最接近的音符是 E5（f_0=659.3Hz）。

在合奏时，如果其他乐器音符的基频接近于这些结合音，就会产生拍音。但是，由于听觉的掩蔽效应，这种拍音现象并不像想象的那样明显。掩蔽效应将在下一节介绍。

5.5.2　掩蔽效应

当我们聆听音乐时，只听到一个纯音的可能性极小。虽然在实验室里或通过电子合成器，可以较容易地听到某个特定频率的纯音（这是有用的、重要的和有价值的体验），然而这种声音不会产生持续的乐感。我们听到的所有乐音至少包含两个频率成分。

当同时聆听两个或更多的纯音时，听觉会产生所谓的“掩蔽”效应，即每个纯音都变得更难听清或不能听清，或者说被另一个声音部分地或完全地“掩蔽”。在这种情况下，我们将产生掩蔽的声音称为“掩蔽声”，

被掩蔽的声音称为“被掩蔽声”。这样的多个声音也可能是由纯音组成。但是在音乐中极少存在这种情况，更多的情况是某个乐器的乐音不同频率成分之间互相掩蔽，或者与另一个乐音的频率成分之间的互相掩蔽。掩蔽的程度取决于掩蔽声和被掩蔽声的频率和幅度。

就像大多数的心理声学研究一样，掩蔽效应的研究也是通过研究某个纯音对其他声音产生的掩蔽效果展开的，并通过了解各个单音的掩蔽特性将结果扩展到复音的掩蔽特性（这和第 3 章的 3.3.2 节中讨论谐和与不谐和的方法类似）。如果用心理声学术语来描述，可以说掩蔽效应是由于掩蔽声的存在使被掩蔽声的听阈改变了，这就为掩蔽效应的定量描述提供了基础，即用掩蔽声存在时听阈曲线的偏移量（也称为掩蔽量）来表示。

掩蔽效应对掩蔽声和被掩蔽声频率的依赖关系可参看图 2.9 所示的理想化听觉滤波器的频率响应曲线。听觉滤波器的带宽是其中心频率对应的临界频带。当输入声音信号所包含的频率成分落在频带内时，滤波器会产生响应，而且由于滤波器响应曲线的不对称性，其对比中心频率低的频率成分的响应要大于比中心频率高的响应。因此，掩蔽效应可以认为是，当掩蔽声落入滤波器频带内时，以被掩蔽声频率为中心频率的这个听觉滤波器对被掩蔽声的有效性（输出增益）在一定程度降低了。滤波器有效性的降低一般用听阈的偏移量或“掩蔽量”表示，如图 5.8(a)所示。图中的曲线显示了滤波器响应曲线的不对称性对掩蔽效应的影响，使得对较高频率的掩蔽效应远大于对较低频率的掩蔽效应，这种现象通常称为“掩蔽的向上延展性”或“低频容易掩蔽高频”。

掩蔽效应与掩蔽声幅度的依赖关系如图 5.8(b)所示，图中表示频率为 f_{masker} 的不同掩蔽声幅度时的理想化掩蔽量曲线。当幅度较低时，掩蔽效应呈现频率 f_{masker} 左右的对称性。随着掩蔽声幅度的提高，低频容易掩蔽高频的效应加强，使得所产生的掩蔽量曲线的不对称性增大。因此，掩蔽效应与掩蔽声的幅度关系较为密切。这种效应可用 Sundberg（1991）的图 5.9 描述。图中的频率刻度用一个临界频带占用等间隔的方式表示。Sundberg 在图中用三组直线近似有掩蔽声时的听阈曲线或“掩蔽阈”曲线：

（1）掩蔽声所在的临界频带之上频段的掩蔽阈以每临界频带降低 5dB ~ 13dB 的斜率下降；

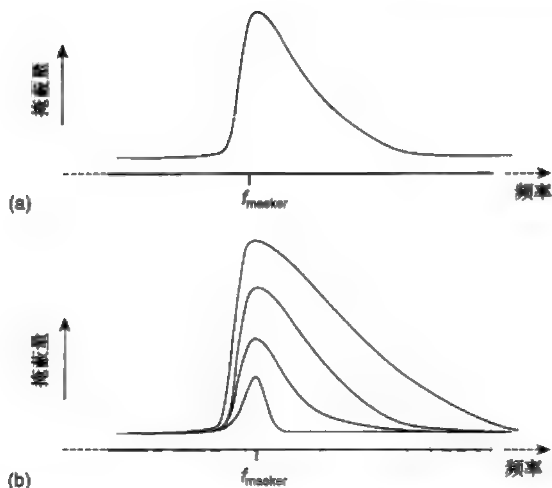


图 5.8　(a) 频率为 f_{masker} 的掩蔽声的"低频容易掩蔽高频"或"掩蔽的向上延展性"的理想化掩蔽量曲线；(b) 掩蔽量随频率为 f_{masker} 的掩蔽声幅度变化曲线

(2) 掩蔽声所在的临界频带的掩蔽阈比掩蔽声的声压级大约低 20dB；

(3) 掩蔽声所在的临界频带之下频段的掩蔽阈变化的斜率比之上的斜率大得多。

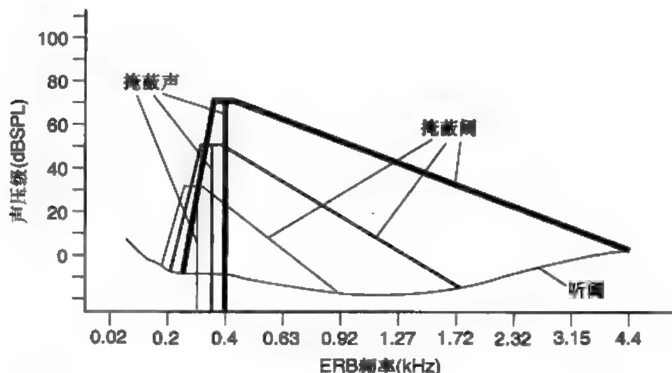


图 5.9　以声压级分别为 50dB、70dB、90dB 的 300Hz、350Hz、400Hz 纯音为掩蔽声的理想化掩蔽阈曲线（频率轴以临界频带刻度，数据源于 Sundberg 1991）

乐音一般是由不同频率成分组成的，其掩蔽效应取决于各频率成分之间的相互掩蔽作用。如果某个频率成分完全被其他频率成分掩蔽，那么被掩蔽的频率成分对音质不会产生任何影响，可以被忽略。如果掩蔽声是像"白噪声"那样的宽带噪声，那么所有频率成分基本上以线性方式被掩蔽（即噪声声压级增加 10dB，则所有频率的掩蔽量增大 10dB）。例如，当存在背景噪声或者用声音频带很宽的响弦鼓鼓刷音（见图 3.6）伴奏时，就属于这种情况。这时落在这些频带里的其他声音的频率成分

就会被掩蔽掉。

目前为止所讨论的掩蔽效应发生在掩蔽声与被掩蔽声同时出现的时候，因此这种掩蔽效应称为“同时掩蔽”。后面将要讨论的对乐音听感非常重要的是掩蔽声和被掩蔽声不同时出现的掩蔽效应，称为“非同时掩蔽”。“非同时掩蔽”分为两种，即前掩蔽和后掩蔽。在前掩蔽中，一个纯音掩蔽声能够掩蔽在它结束以后出现的另一个声音（被掩蔽声）。换句话说，掩蔽声出现在被掩蔽声之前。前掩蔽发生的有效时间间隔，即掩蔽声结束到被掩蔽声开始的最大时间间隔可达 30ms。在后掩蔽中，一个声音可以被在它之后出现的掩蔽声所掩蔽，最大有效时间间隔大约是 10ms。应该注意到，前掩蔽和后掩蔽发生的最大有效时间间隔因人而异。

同时掩蔽和非同时掩蔽效应可用图 5.10 所示的掩蔽量理想化时域曲线来总结。图的下部给出了掩蔽声开始和结束的时间段，并假设同时掩蔽的掩蔽量为 50dB。图中也显示了后掩蔽和前掩蔽效应随时间间隔变化的特性。Moore（1996）对非同时掩蔽效应作出以下总结：

（1）实际上后掩蔽效应非常小（在某些情况下几乎不存在）；

（2）掩蔽声的声压级越高，前掩蔽曲线的变化速率越大；

（3）掩蔽声停止 100ms ～ 200ms 后，前掩蔽效应消失；

（4）当掩蔽声持续时间在 50ms 以内时，前掩蔽效应随掩蔽声持续时间的增大而增大。

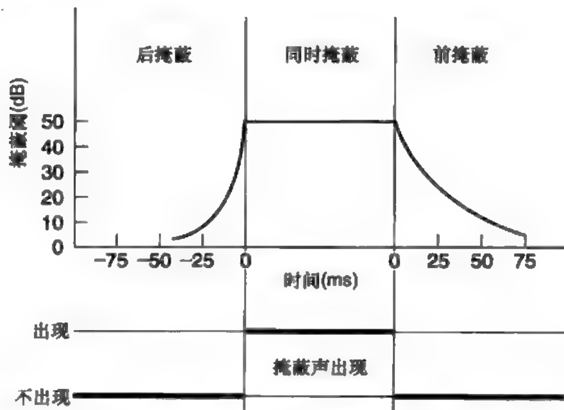


图 5.10　同时掩蔽和非同时掩蔽的理想化特性曲线

在储存和传输数字声频信号的系统中，掩蔽效应用于减小所要处理的信息量，从而减小所需的传输资源或传输带宽，并提高存储器、光盘

或其他存储媒介的存储效率。由于这些系统利用了人耳听觉特性，因此称为感知编码系统。例如，MP3 系统就是以感知编码为基础的。感知编码还应用于因特网的音乐传输、可以存储好几个小时音乐的小型 MP3 播放器、数字声频广播的多声道环绕声、小型激光唱片录音机（Maes 1996）和现在已经过时的数字小型盒带机（DCC）。

感知编码系统包括编码（用于制作阶段）和解码（用于播放阶段），它们需要国际标准规定系统的基本运行策略，以便不同厂家从事研究和生产。移动图像专家组（MPEG）成立于 1988 年，他们的任务一直都是为移动图像和所伴随的声音的编码制定国际标准。已建立的标准有 MPEG-1、MPEG-2 和 MPEG-4，每一种都包含三层：层一、层二和层三。MP3 就是 MPEG-1 的第三层算法（千万不要认为是 MPEG-3！）。声音的感知编码原理将在下面简述，详细资料参看参考文献 Watkinson（1994、1999）、Gilchrist 和 Grewin（1996）以及 Rumsey（1996）。

图 5.11 所示为音频感知编码信号处理的基本方框图。输入信号首先通过一系列带通滤波器分解为不同频带的信号（方框 1）。这些带通滤波器有时称为“子带”滤波器，因此这类编码器常被称为“子带”编码器。这个处理过程与听觉外围系统的临界频带的吻合程度取决于特定编码算法本身的复杂性。然后每个子带信号能量与原输入信号进行对比，用于计算该时刻的同时掩蔽效应（某些情况下也包括非同时掩蔽效应）（方框 2），那些系统判定不会被掩蔽的子带信号经过数字编码（方框 3）后用于传输和存储。在接收端，具有相反的信号处理过程的解码器将信号还原出来，当然所还原的信号不可能与原信号完全相同，因为那些听觉感受不到的声音信息已经被去除了。

关于听觉掩蔽效应的听音演示可以从 Houtsma 等人（1987）的 CD 获得。

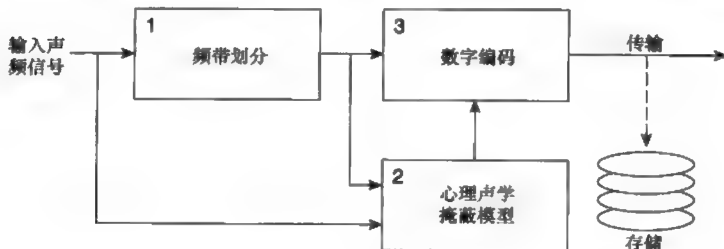


图 5.11　感知编码系统基本方框图

5.5.3　乐音的编组幻觉

在某些情况下，听觉听到的声音是人们没有预料到的，这是听觉产生幻觉的结果，这种现象与听觉系统分析声音信号的方式有关。有些这样的听音信号只能用电子合成的方式产生，用声学乐器演奏的传统音乐一般不会产生这样的听觉效果，而有些这样的效应与音乐的表现方式有一定的关系。了解听觉幻觉的特性及其与输入声音信号之间的关系，有助于在理论上进一步提升对人耳如何感知声音这个问题的认识。有些效应可能已经应用于音乐演出中，或者可以期望其在将来的音乐演出中发挥作用。

Diana Deutsch 对声音编组幻觉进行了研究，下面对其中一些结果进行概述，说明其在音乐听音和表演中的表现。Deutsch（1974）描述了一种"八度幻觉"，即将两个音高相差八度的（较高基频为 800Hz、较低基频为 400Hz）乐音，按图 5.12 上图所示的方式交替馈送到听音者两耳。无论耳机以何种方式馈送信号，大多数听音者的反应是交替地在右耳听到高音而在左耳听到低音，如图所示。她进一步指出，右撇子（多数人）听音者往往在右耳听到高音而在左耳听到低音，但是左撇子听音者在右耳和左耳听到高音的概率相等。当用扬声器重放激励声时，这种幻觉同样存在。用于这项听音演示的激励声可以在 Houtsma 等人（1987）的 CD 中得到。

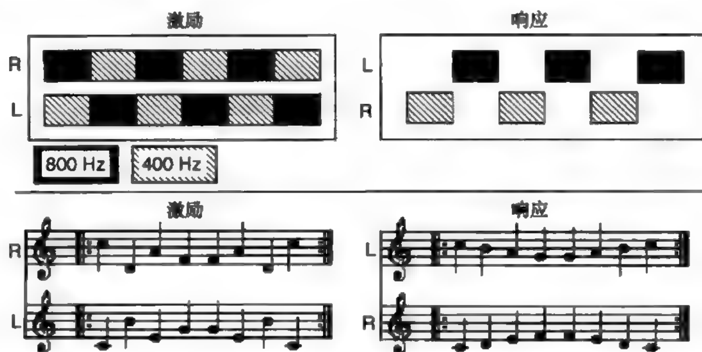


图 5.12　Deutsch（1974）"八度幻觉"的激励和响应（上）以及 Deutsch（1975）"音阶幻觉"的图示（下）

在下一个实验中，Deutsch（1975）将 C 大调上升音阶和下降音阶分别交替地馈送给两耳，如图 5.12 下图所示。最常见的响应如图所示，高

音再次倾向于被右耳听到，左耳听到的是较低的音符，左右耳分别听到 C 大调音阶的片断，这种现象称为“编组”。Deutsch 用心理学中称为“连续”、“相近”和“相似”的“格式塔学派”的编组原理解释这种现象，阐述如下（Deutsch 1982）：

（1）连续编组——“在某一特定方向或以某种方式互相跟随的元素被编组在一起”；

（2）相近编组——“较接近的元素优先于相距较远的元素被编组在一起”；
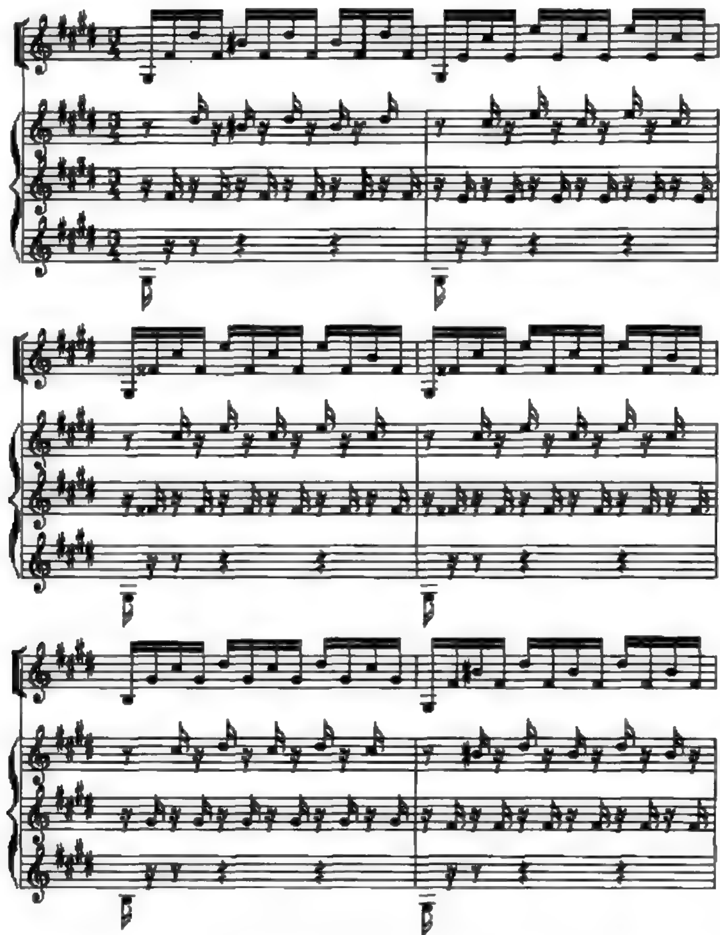
（3）相似编组——“相似的元素被编组在一起”。

在各种情况下，“元素”均指激励中的各个音符或乐音。当把这些原理应用于图中所示的激励时，Deutsch 指出，将高音音符或低音音符编组的相近编组效应要比可能会感觉到 C 大调的完整上升或下降音阶的连续编组效应更显著。Deutsch（1982）用其他的一些实验支持她的这个观点。尽管只有一条旋律，由于聆听音乐时听觉的频率相近编组效应，使得听觉感受到多条旋律同时在进行。产生这种效应的独奏音乐作品包括 J.S.Bach 的小提琴组曲和奏鸣曲。图 5.13 所示为 J.S.Bach 的小提琴组曲 III 的 E 大调前奏曲所产生的编组效应示意图（本书 CD 的第 66 条），图中显示出乐谱（最上面一条五线谱）和听觉通常感受到的 3 个部分旋律（下面 3 条五线谱）。可见听觉是按频率相近原理编组的。

柴可夫斯基第六交响乐最后一个乐章的非常不寻常的弦乐片段（本书 CD 的第 67 条）也常常在讨论听觉编组效应时被提到。听觉通常会感觉到 4 个部分的旋律，如图 5.14 所示。这种现象同样可以用听觉的频率相近编组原理来解释。假设乐队的摆位是按当时的方式，即弦乐从左到右的顺序是：第一小提琴、低音提琴、大提琴、中提琴和第二小提琴，听觉编组产生的效果可以用聆听立体声来比喻。但这种摆位方式和现在普遍采用的正好相反（从左到右）：第一小提琴、第二小提琴、中提琴、低音提琴和大提琴。

其他听音幻觉可能还会由音色相近编组产生。Pierce（1992）阐述了“由 David L.Wessel 在 1978 年描述”的一个实验，如图 5.15 所示。在这个实验中，当所有音符的音色相同时，激励（A）在听觉产生的响应是上升的琶音，即响应（A）；然而，当音符的音色在两种音色之间交替变化（图中

图 5.13 J.S.Bach 的小提琴组曲 Ⅲ 的 E 大调前奏曲的第 45 ~ 50 小节的乐谱（上面第一条五线谱）和听觉编组产生的 3 个部分旋律（下面 3 条五线谱）



用不同形状的符头表示），且音色差别较大时，如图中激励（B）所示，听觉会听到两组下降的琶音，如图中响应（B）所示。这种现象称为音色编组效应（Bregman 1990）。图 5.16 所示的音符由 4 种不同形状符头表示的不同音色组成，用来说明音色编组效应的可变性（这是音乐技术学生的实验室练习之一）。乐谱可以无限重复下去，并且速度可以变化，听觉听到的是上升的音阶还是下降的音阶主要取决于演奏这一系列音符的速度。当速度较低（小于每秒一个音符）时，听到的是上升的音阶（图中五线谱 B），编组是以音符出现的顺序为依据的；当演奏速度提高时，例如大于每秒 10 个音符，听觉听到的是不同音色的下降音阶（图中五线谱

C ~ F），这时编组是以音色为依据。听音者通过调整对激励中不同音色的注意力，听觉所感受的声音能够在五线谱 C ~ F 之间切换。
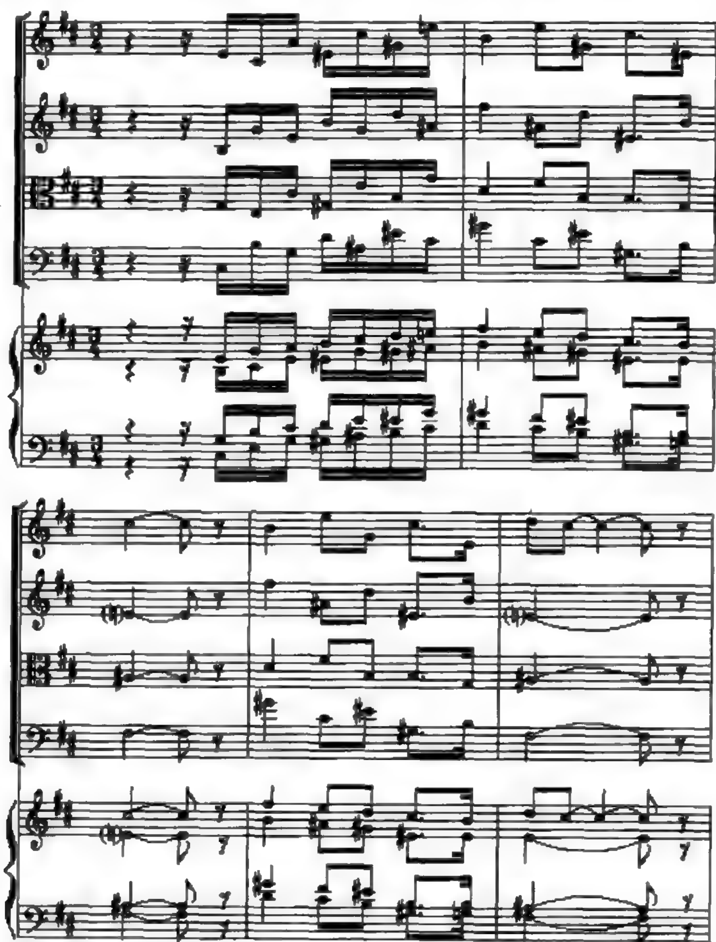


图 5.14　柴可夫斯基第六交响乐的最后一个乐章弦乐片断的乐谱（上）和通常感受到的 4 个部分旋律（下）

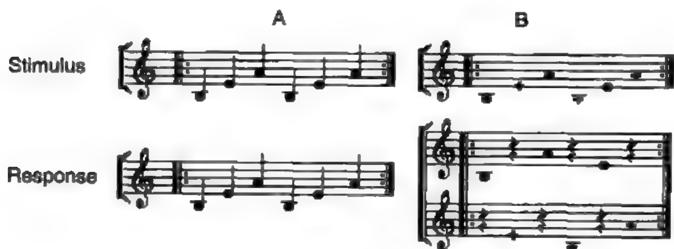


图 5.15

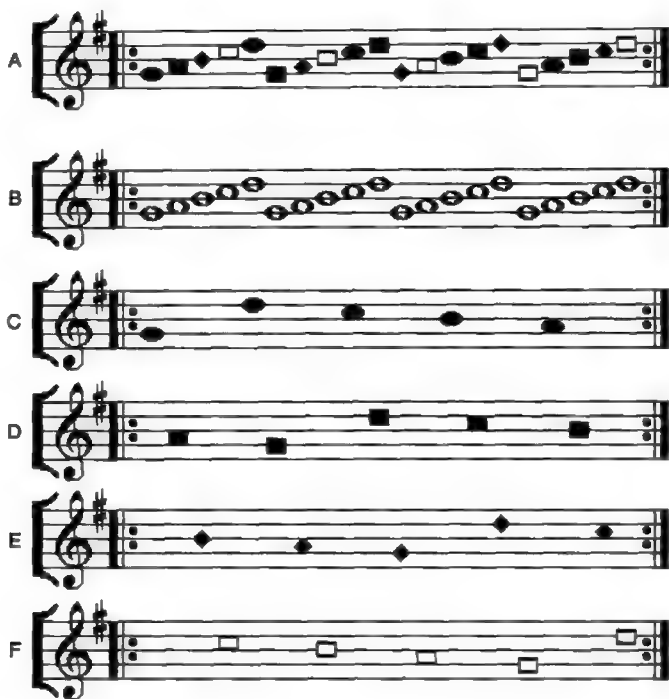
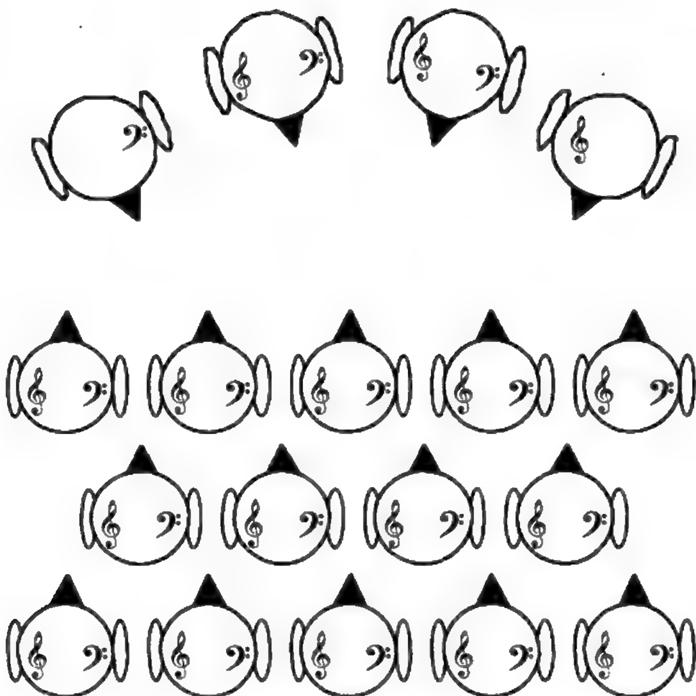


图 5.16 用于音符顺序编组和音色编组实验中的激励（五线谱 A，其中不同符头形状表示不同音色。当演奏速度低时，听觉按音符顺序编组（五线谱 B）；当速度高时，按音色编组（五线谱 C – F）。）

大多数听音者对图 5.12 所示激励的响应，倾向于在右耳听到高音而在左耳听到低音，这种结果与乐队演奏者在舞台的摆位方式有一定关系。例如，弦乐四重奏的摆位是大提琴手位于中提琴手的左边，中提琴手位于第二小提琴手的左边，而第二小提琴手位于第一小提琴手的左边，如图 5.17 所示。这意味着每个演奏者的左手边是较低音高的乐器，而在右手边是较高音高的乐器。合唱队的摆位方式一般是女高音位于女中音的右边，男高音位于男低音的右边（如果排成两行或两行以上时）。小型合唱队如四重唱往往排成一行，由女高音、女中音、男高音和男低音组成，一般男低音位于左边，女高音位于右边。在乐队中，高音乐器（如第一小提琴、短笛、小号等）总是位于相应声部的左侧，而低音乐器位于右侧（译者注：从台下看）。这种摆位方式已经形成一种传统习惯，演奏者或歌唱者不愿意改变他们的相对位置。传统的乐队摆位方式很可能部分地源于听音者的右耳更习惯于倾听高音的听觉特性。

图 5.17　音乐厅中传统的演员和观众的摆位以及演员和观众的双耳高、低音偏好



然而，虽然这种乐队摆位对演员来说是符合听觉习惯的，但对听众来说正好相反。当听众面对舞台观看实况演出时（见图 5.17），高音乐器或高音歌唱者位于舞台左侧，低音乐器或低音歌唱者位于舞台右侧，这与右耳倾向于听到高音的听觉特性相违背，但是对演员自己来说是正确的摆位。可以进行一项有趣的实验，让听音者坐在乐队后方聆听演出，并与前方位置的听音效果进行比较，如伦敦皇家节日音乐厅就在舞台后方设有座位。遗憾的是舞台后方的座位往往距离乐队太近，很难获得与前方观众区同样的总体平衡感，因此，这样的比较实验很难进行。但是，还有一种能够体验不同听音效果的方法是，转过身去聆听扬声器重放的立体声录音或将左右扬声器位置互换。

5.5.4　音高幻觉

音高幻觉由 Shepherd（1964）提出并进行了演示，这种现象可以与 Maurits Escher 的连续楼梯图片相对比，经常称为“Shepherd 音（Shepherd

tone）"。这种幻觉产生的是连绵不断的以半音步长上升的音阶的感觉。因为当音高上升一个八度或者上升 12 个半音后，频谱与开始时的频谱相同，因此，虽然声学频谱上升没有超过一个倍频程，听觉却感觉音阶不断上升。这个声激励可以从 Houtsma 等人（1987）录制的 CD 得到（或本书 CD 的第 68 和 69 条）。图 5.18 所示为产生"Shepherd 音"的声激励频谱特性，图中所示激励只包含基频和基频以上八度整数倍的谐波分量。"Shepherd 音"各频率分量可表示为：

$$f_{shepherd}=(2^n f_0) \quad (5.4)$$

其中，$f_{Shepherd}$ 为"Shepherd 音"各分量的频率，$n=0，1，2，3\cdots$。

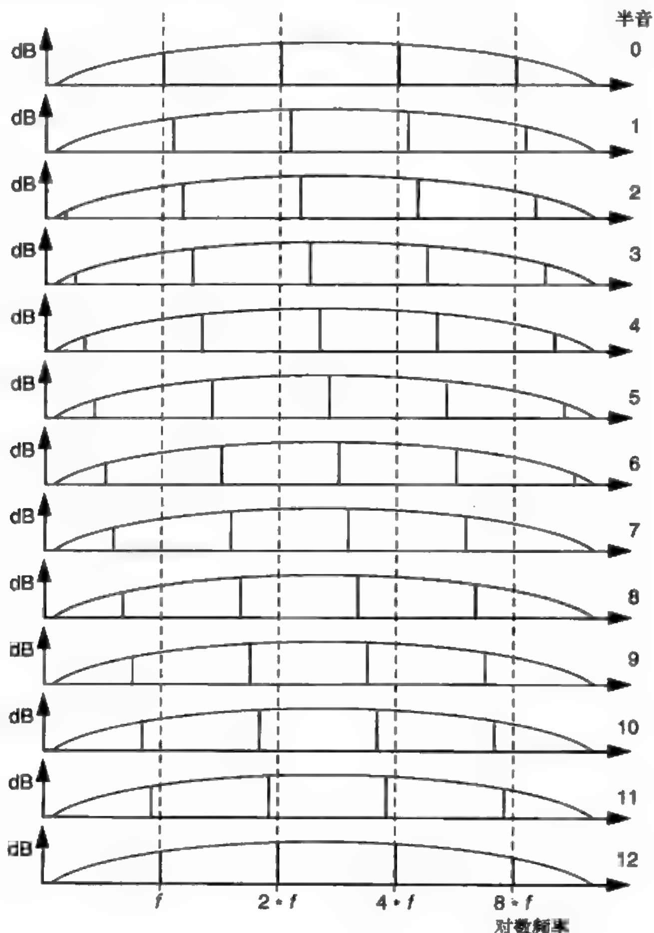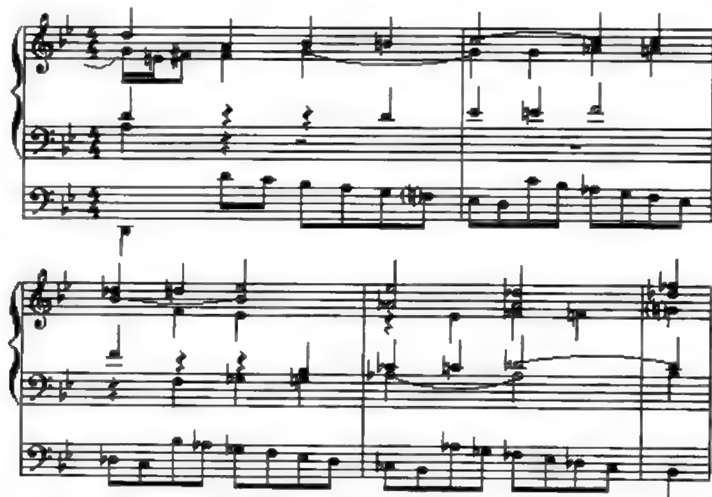


图 5.18 产生 Shepherd 连续上升音阶幻觉的声激励频谱示意图

各频率分量的幅度由图中所示的包络决定。每当音符提高一个半音，频率分量均向上偏移 1/12 倍频程，即一个半音。较高的谐波成分越来越弱并最终消失，新的较低的谐波分量出现并越来越强。

与“Shepherd 音”效应相关的音乐例子为 J.S.Bach 的管风琴曲《G 小调幻想曲》（BWV 542）踏板的第 31 小节以后的段落（本书 CD 的第 70 ～ 72 条），一些听音者能够听到个别模糊的音高。这些小节的管风琴乐谱如图 5.19 所示，上面两条谱线供键盘上左右手使用，下面一条谱线供踏板使用。踏板乐谱由 5 组下降音阶构成，除了最后 1 组外，每组由 8 个音符组成。每组音阶都以跳跃小七度结束。即使眼睛看着乐谱，听音者往往在此处不能明确指出音高变化的时刻。这种效应的显著程度取决于用什么音栓进行演奏。对大多数听音者来说，这种音高感觉的模糊性在所摘录乐谱的第 3 小节比较明显，往往感觉音高的跳跃发生在一个或两个音符之后。原因可能是当踏板部分跳跃到乐谱上的 B♭3 时，左手新增加的音符 F3 刚好开始（这里强调“乐谱上的”B♭3，是因为踏板演奏的实际基频比乐谱表示的基频低八度，参看 5.4 节）。这两个音符 F3 和 B♭2 的基频 f_0 正好是 B♭1 的三次和二次谐波频率，而 B♭1 又是踏板下降音阶在不跳跃一个八度时的下一个音符。在所有踏板下降音阶的跳跃过程中，键盘上的和声从小调变化为大调。在第 32 小节中，左手从 E♭变化到自然 E，增加了新的谐波成分，其中包含踏板下降音阶在不跳跃一个八度时

图 5.19　J.S.Bach 的管风琴曲《G 小调幻想曲》（BWV 542）第 31 小节以后的片断



的下一个音符（乐谱上的 C3），而 E♭不是这个谐波系列的成员。同样地，在所摘录乐谱的第 3 小节，右手自然 D 加入左手 F3，在第 4 小节加入了左手自然 C。这些加入的音符和踏板下降音阶在不跳跃一个八度时的下一个音符的基频属于同一个谐波系列，这是听觉不能明确音高向上跳跃的准确时间的原因。

在管风琴的踏板合成低音音栓中，也应用了听觉幻觉特性，即通过音管的组合产生比实际存在的任何一个音符都低的音高感觉。所产生的音高为 32'（极少情况下为 64'），它们最低 C 的基频 f_0 分别是 16.25Hz 和 8.175Hz。一个 32' 音高的合成低音音栓是通过组合 16' 音栓和 10⅔' 音栓形成的，这两个音栓的基频正好是 32' 谐波系列的二次和三次谐波（参看 5.4 节）。一个可能标有“声学低音（acoustic s bass）”字样的 32' 音栓其实是一个 10⅔' 的变化音栓，当它与 16' 音栓共同发声时，会产生 32' 的音高感觉（见第 3 章关于从二次和三次谐波产生音高感觉的部位学说）。类似地，标有“合成低音（resultant bass）”的 64' 音栓是由一组 21⅓' 音管和一组 32' 音管共同发声产生的。32' 音栓的中央 C（C2）基频 f_0 为 65.4Hz，其最低 C（C0）比中央 C 低两个八度，基频 f_0 为 65.4 除以 4 即 16.35Hz。64' 音栓的最低 C（C-1）的基频为 8.175Hz，这个频率在听觉能够感知的最低频率之下，但是在听觉产生拍音的频率差之内（见图 2.6）。在听觉频率之内的谐波成分能够产生 f_0 在听觉频率之外的音高感觉。管风琴演奏者有时通过在踏板上演奏五度和声来模拟这种声音效果，尤其经常在乐曲的最后一个音符中使用，但是结果并不像由真正合成低音音栓产生的声音那样令人满意，因为最好的效果要求三次谐波（10⅔'）比二次谐波（16'）弱。

Roederer（1975）以管风琴乐曲为例描述了一种称为“残留音高”的听觉现象，它属于音高幻觉的一种，听音演示见本书 CD 的第 74 条，而第 73 条为通常听到的众赞歌。他选用 J.S.Bach 的管风琴曲集的众赞歌第 40 号作品。众赞歌前奏曲的独奏部分由若干个变化音栓组合演奏（如 8'、4'、2⅔'、2'、1⅗'、1⅓'、1'），并伴随左手演奏 8'、4'，踏板演奏 16'、8'。对一个受过听音训练的听音者的要求是，跟踪旋律音高变化，并且告知可能发生音色变化。当完成一个片断演奏后，继续演奏，但去掉 8'，然后去掉 4'，接着去掉 2'，最后去掉 1'，最终独奏部分只包含变化音栓（即

那些与 8' 基频关系不为 1 或整数个八度关系的音栓)。Roederer 建议："让听音者注意到高音部分能听到什么，并告诉他们所有乐谱上记录的音高都缺失（包括所有八度关系的谐波成分）——他们会感到难以置信！这样的实验是不可重复的，因为在第二次实验时听音者的听觉系统会更改音高处理策略！" 经验表明，这种听音实验依赖于这类激励声所建立的音高关系，通常采用熟悉的或连续的音乐旋律。

乐音幻觉只有在听觉思维存在强烈期望的前提下才可能发生。

参考文献

ANSI (1960). American Standard Acoustical Terminology. New York: American National Standards Institute.

Bregman, A.S. (1990). Auditory Scene Analysis. Cambridge: MIT Press.

Brookes, T., Tyrrell, A.M., and Howard, D.M. (2000). On the differences between conventional and auditory spectrograms of English consonants. Logopedics, Phoniatrics, Vocology, 25, 72–78.

Deutsch, D. (1974). An auditory illusion. Nature, 251, 307–309.

Deutsch, D. (1975). Musical illusions. Scientific American, 233, 92–104.

Deutsch, D. (1982). Grouping mechanisms in music. In The Psychology of Music (J. Deutsch, ed.). London: Academic Press.

Deutsch, D. (1983). Journal of the Audio Engineering Society, 31, (9), 607–622.

Gilchrist, N. and Grewin, C. (eds) (1996). Collected Papers on Digital Audio Bit-rate Reduction. The Audio Engineering Society.

Grey, J. (1977). Timbre discrimination in musical patterns. Journal of the Acoustical Society of America, 64, 467–472.

Hall, D.E. (1991). Musical Acoustics: An introduction (2nd edn). Belmont, CA: Wadsworth Publishing Company.

Helmholtz, H.L.F. von (1877). On the Sensations of Tone as a Physiological Basis for the Theory of Music (4th edn) trans. Ellis, A.J. New York: Dover (1954).

Houtsma, A.J.M., Rossing, T.D., and Wagenaars, W.M. (1987). Auditory

Demonstrations. (Philips compact disc No.1126–061 and text.) New York: Acoustical Society of America.

Howard, D.M., Hirson, A., Brookes, T., and Tyrrell, A.M. (1995). Spectrography of disputed speech samples by peripheral human hearing modelling. Forensic Linguistics, 2, (1), 28–38.

Howard, D.M. and Tyrrell, A.M. (1997). Psychoacoustically informed spectrography and timbre. Organised Sound, 2, (2), 65–76. Hearing timbre and deceiving the ear.

Howard, D.M. (2005). Human hearing modelling real-time spectrography for visual feedback in singing training. Folia Phoniatrica et Logopaedica, 57, (5–6), 328–341.

Maes, J. (1996). The Minidisk. Oxford: Focal Press.

Moore, B.C.J. (1996). Masking and the human auditory system, in Collected Papers on Digital Audio Bit-rate Reduction (N. Gilchrist and C. Grewin, eds). The Audio Engineering Society, 9–22.

Padgham, C.A. (1986). The Well-tempered Organ. Oxford: Positif Press.

Pierce, J.R. (1992). The Science of Musical Sound. Scientific American Books (2nd edn). New York: W.H. Freeman and Company.

Pollard, H.F. and Jansson, E.V. (1982). A tristimulus method for the specification of musical timbre. Acustica, 51, 162–171.

Rasch, R.A. and Plomp, R. (1982). The perception of musical tones, in The Psychology of Music (D. Deutsch, ed.). London: Academic Press.

Roederer, J.G. (1975). Introduction to the Physics and Psychophysics of Music. New York: Springer-Verlag.

Rumsey, F. (1996). The Audio Workstation Handbook. Oxford: Focal Press.

Scholes, P.A. (1970). The Oxford Companion to Music. London: Oxford University Press.

Shepherd, R.N. (1964). Circularity in pitch judgement. Journal of the Acoustical Society of America, 36, 2346–2353.

Sumner, W.L. (1975). The Organ. London: MacDonald and Company.

Sundberg, J. (1991). The Science of Musical Sounds. San Diego: Academic

Press.

Watkinson, J. (1994). The Art of Digital Audio (2nd edn). Oxford: Focal Press.

Watkinson, J. (1999). MPEG 2. Oxford: Focal Press.

第 6 章　听音的声学环境

本章将介绍声音如何在房间里传播，重点介绍房间声学特性对听音的影响。同时，我们也探究分析和改善房间音质的方法。最后，我们将了解听觉怎样判定声音的方向，在此基础上讨论如何进行听音室、录音棚和控制室的声学设计，以创造良好的立体声听音环境。

6.1　封闭空间的声学特性

第 1 章主要讨论了在无边界的环境下声波的传播特性。然而，大多数音乐是在房间里聆听的，受到房间边界的影响。因此，了解声音如何在封闭空间里传播是非常重要的。图 6.1 所示为一个理想化房间，房间里

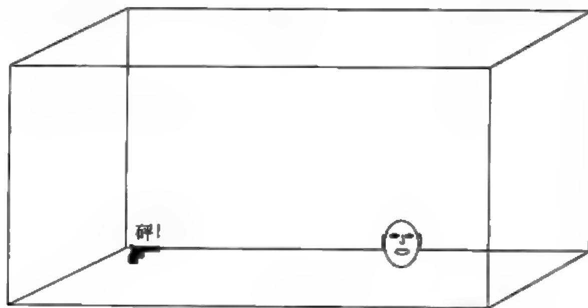


图 6.1　用手枪发出脉冲声激励的理想化房间

有一支手枪和一个听音者。假设手枪在某一时刻（t=0）发声，下面从 3 个主要方面讨论手枪的声音在房间里的传播情况。

6.1.1 直达声

手枪发声后一瞬间，听音者将听到手枪发出的声音，这个声音从手枪直接到达听音者，传播距离最短，延迟时间可以用距离计算，声音的传播速度是每秒 344m。由于从声源到听音者的最短距离为最直接的距离，因此听音者首先听到的是这个声音。这个声音分量称为直达声，它的传播路径和相应的时间响应如图 6.2 所示。

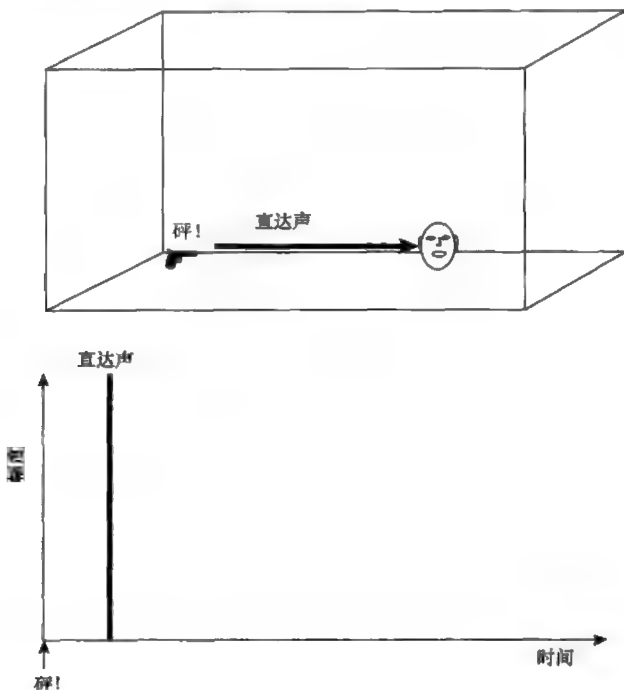


图 6.2 房间的直达声

由于直达声携带的声音信号不受周围环境影响，因此这个声音十分重要。为了获得清晰的声音和好的语言清晰度，要求直达声具有较高的声压级。直达声的传播特性与自由空间的传播特性相同，因为它没有和边界发生相互作用，因此我们可以用自由空间声强计算公式来计算离声源一定距离的直达声的声强。由第 1 章可知，直达声的声强为：

$$I_d = \frac{QW_{Source}}{4\pi r^2} \tag{6.1}$$

其中，I_d 为声强（单位是 W/m²），Q 为声源的指向性因数（以球形声源为参考），r 为到声源的距离（m）。

式（6.1）说明，直达声的声强与距离的平方成反比，完全与声音在自由空间的情况相同，并且对实际空间的听音有较大影响。下面举例计算扬声器发出的直达声声强。

例 6.1 某扬声器在 1m 处产生的声强级为 102dB，试问距扬声器 4m 处的直达声声强级 I_d 是多少?

利用第 1 章的式（1.18），可计算一定距离的直达声声强级为：

$$SIL = 10\lg\left(\frac{W_{Source}}{W_{ref}}\right) - 20\lg(r) - 11\text{dB}$$

由于已知 1m 处的声强级，这个公式变形为：

$$I_d = I_{1m} - 20\lg(r)$$

因此，直达声声强级为：

$$I_d = 102 - 20\lg4 = 102 - 12 = 90\text{dB}$$

这个例子说明距离对直达声声强级的影响十分显著。

6.1.2 早期反射声

听到直达声稍后一段时间内，听音者会听到经过一个或更多表面（墙壁、地板等）反射后的声音，如图 6.3 所示。这些声音称为早期反射声，它们在到达时间和到达方向上都与直达声不同，人们可以利用这种不同来判断房间的大小和声源的空间位置。如果这些反射声相对直达声的延迟时间较长，当总声程差大于约 30ms（10.2m）时，人们可能会听到回声。早期反射声会引起干涉现象（见第 1 章），使房间里语声的清晰度减小，并对音乐产生人们所不希望的音色变化。早期反射声的声强级与传播距离以及反射面有关。通常表面会吸收一定的声能，因此声吸收会使反射声强度进一步变弱。但是，有些表面会产生声聚焦，使听音者所在位置的声音加强，如图 6.4 所示。尽管会出现这种情况，但是由于表面的吸声作用，声

音的总能量还是减少了，这一点需要注意。声聚焦虽然使某些位置的声音加强了，它却使其他位置的声音强度降低了。产生声聚焦的表面尺寸应该比声波波长大，这就是说高频比低频更容易产生声聚焦。一般来说，由于表面的吸声作用，反射声的声压级要比仅仅依据距离平方反比定律的直达声的声压级小。下面举例计算扬声器的早期反射声强度。

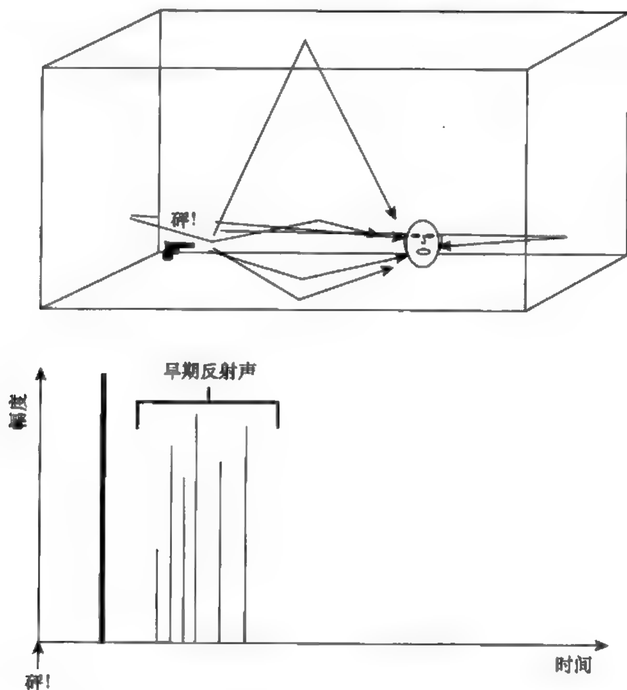


图 6.3　房间的早期反射声

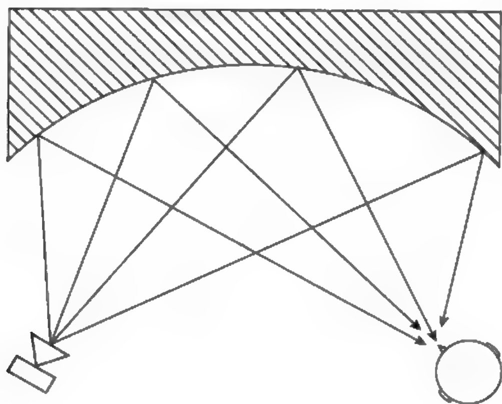


图 6.4　声聚焦面

例 6.2 某扬声器在 1m 处产生的最大声强级为 102dB。当扬声器距离坚硬反射墙 1.5m，并且听音者站在扬声器前方 4m 处时，试问其产生的反射声的声强级 I_r 和相对于直达声的延迟时间是多少？

题图如图 6.5 所示。图中给出扬声器相对墙面产生的镜像声源，可以根据"毕达哥拉斯"定理计算出声音由于反射产生的附加路程。计算后得出反射声的路程 L 为 5m。

由已知的 1m 处的声强级，并根据平方反比定律，可计算出反射声的声强级为：

$$I_r=I_{1m}-20\lg L \tag{6.2}$$

$$I_r=102-20\lg 5=102-14=88\text{dB}$$

与上一个例题比较后可知，反射声的声强级比直达声低 2dB。延迟时间可以简单地从传播路程计算如下：

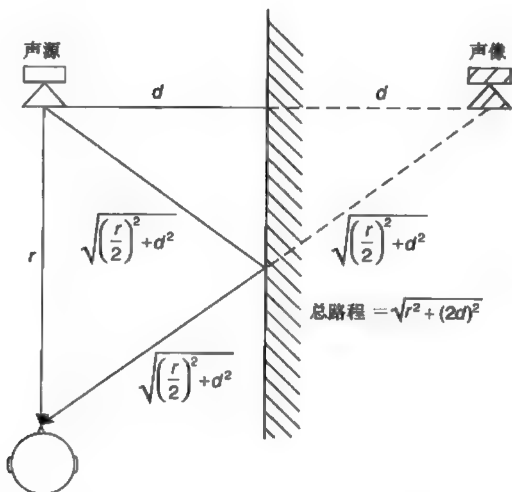
$$\frac{5}{344}=14.5\text{ms}$$

同理，直达声的延迟时间为：

$$\frac{4}{344}=11.6\text{ms}$$

因此，反射声相对于直达声到达听音者的延迟时间为 14.5ms−11.6ms = 2.9ms。

图 6.5 计算早期反射声强度的几何图形



由声强级的平方反比定律可知，声音的强度与延迟时间、到声源的距离之间存在直接的关系，我们可以把这种关系用图表示出来。图 6.6 所示为当不考虑吸声或声聚焦效应时反射声声强级的最大值。

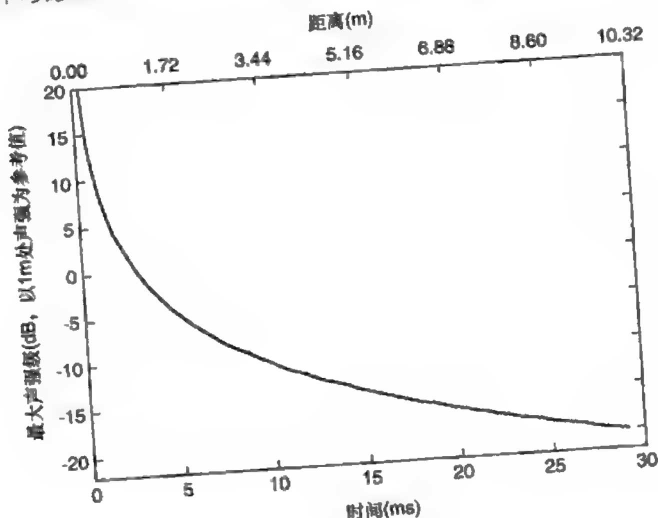


图 6.6　早期反射声声强级上限（假设不存在声吸收和声聚焦）

6.1.3　吸声对早期反射声的影响

声吸收如何影响早期反射声的声压级呢？材料的吸声系数决定了当声波碰撞它时有多少声能或声功率被吸收掉。一般来说，实际材料的吸声系数随频率变化，但是，在这里我们首先假设它不随频率变化。一定面积的吸声材料所吸收的声能或声功率与入射声波的单位面积声能或声功率有关。由于声强是由单位面积声功率决定的，因此可以认为反射声的声强的减小与吸声系数成正比，即：

$$I_r = I_i \times (1-\alpha) \qquad (6.3)$$

其中，I_r 为反射声声强（W/m^2），I_i 为入射声声强（W/m^2），α 为吸声系数。

由于声强的乘积运算相当于声强级的加法运算，因此式（6.3）可以用声强级表示如下：

$$I_r = I_i + 10\lg(1-\alpha) \qquad (6.4)$$

上式与式（6.2）结合后，得到经过吸声表面反射后早期反射声声强级计

算公式为：

$$I_r=I_{1m}-20\lg L+10\lg(1-\alpha) \qquad (6.5)$$

下面通过举例计算，说明吸声表面对早期反射声强度的影响，并与前面例子的计算结果进行比较。

例 6.3　某扬声器的最大声强级为 1m 处 102dB。当扬声器距离反射墙 1.5m，并且听音者站在扬声器前方 4m 处时，试问其产生的反射声声强级（I_r）是多少？设墙面的吸声系数分别为 0.9、0.69 和 0.5。

因为已知 1m 处的声强级，并且反射声同样遵守平方反比定律，因此早期反射声声强级可用式（6.5）计算：

$$I_r=I_{1m}-20\lg L+10\lg(1-\alpha)$$

前面例题已经计算出反射声路程为 5m，因此，3 种不同吸声系数时听音者处的声强级为：

$$I_r=102-20\lg 5+10\lg(1-0.9)=102-14-10=78\text{dB}$$

$$I_r=88+10\lg(1-0.69)=88-5=83\text{dB}$$

$$I_r=88+10\lg(1-0.5)=88-3=85\text{dB}$$

6.1.4　混响声

听到直达声的更长时间后，声音经过多次反射并从各个方向到达听音者，如图 6.7 所示。由于声音从如此多不同的方向和路径到达听音者，因此各个反射声彼此在到达时间上非常靠近，听音者听到的是一组密集的反射声，这部分声音称为混响声。混响声具有丰富和烘托音乐的作用，因此是人们所希望听到的。混响声还能使不同乐器的声音融合，使听音者听到相互融合的所有乐器的声音。实际上我们会发现在没有混响的房间听音很不舒服，一般不喜欢在这样的房间里演奏音乐，而在浴室里唱歌要比在客厅里唱歌有趣得多（参考本书 CD 的第 75 ~ 77 条）。混响时间长短与房间的大小有关。由于两次反射之间的时间间隔较短，而每次反射都存在能量损失，所以房间越小，混响越短。事实上，直达声与混响声之间的时间间隔有助于听觉判断音乐演奏所在空间的大小。由于每次反射都有部分能量被吸声，因此混响最终会消失。混响消失所需的时

间称为混响时间，它与房间大小以及每次反射的吸声量有关。混响声场可分为 3 个阶段，如图 6.8 所示。

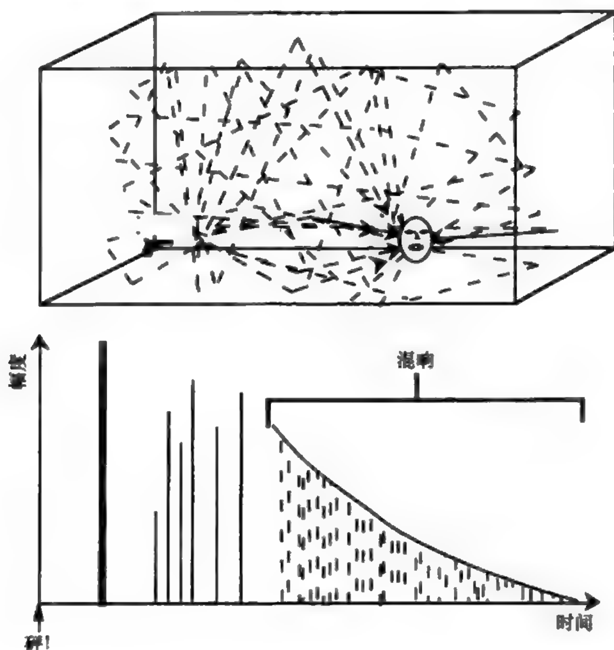


图 6.7　房间的混响声

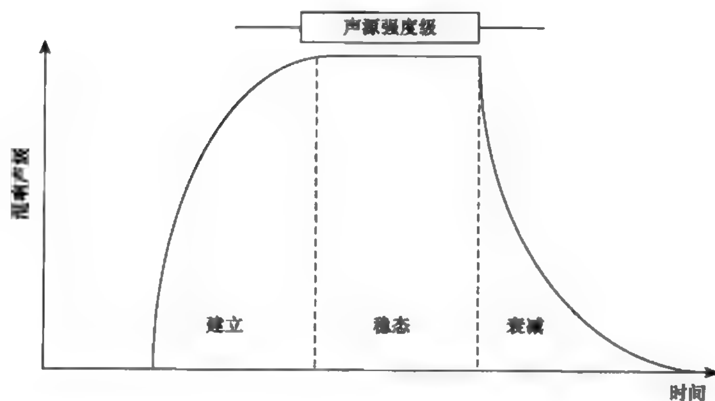


图 6.8　房间里混响声级随时间变化过程

（1）混响声场的建立：这是混响声场的最初部分，与房间的尺寸有关。因为房间尺寸影响两次反射的时间间隔，从而影响建立混响声场所需的

时间。房间的吸声量也会影响稳态混响声场的建立时间，而且稳态混响声压级与房间的吸声量成反比，这在后面会介绍。由于声场建立的速率与两次反射或吸收的时间间隔有关，因此较大的房间比较小的房间需要更多时间来建立混响声场。

（2）稳态混响声场：当在房间里演奏某个稳定的乐音如管风琴乐音时，一定时间以后，输入的声功率会等于由于吸声而损失的声功率，这时混响声场达到一个稳定的声压级。这就是说，较小吸声量的房间比较大吸声量的房间具有更高的稳态混响声压级。但是，一个瞬态声在房间里不能达到稳态声压级。

（3）混响声场的衰减：在某个乐音停止的瞬间，房间里的混响声压级不会马上变为零，而是以一定的速率衰减。衰减速率取决于每次反射吸声量的大小，因此在吸声量较小的房间，混响声场需要更多的时间完成衰减过程。

较大的房间会产生较长的混响时间，房间里摆设的家具等物品越多，混响时间越短。从较小的、摆满家具的房间到宽大的、有许多玻璃窗和石头墙的大教堂，混响时间可能在 0.2s ～ 10s 变化。

6.1.5 混响声场的特性

从听音者的角度看，混响声与直达声和早期反射声有很大不同。直达声和早期反射声都满足平方反比定律，此外早期反射声还存在吸声效应，因此它们的声压级随位置变化。而房间里的混响声大小几乎不随听音者位置变化。这并不是因为混响声与一般声波的表现不一样，而是因为混响声是从各个不同方向到达听音者，使得在房间不同位置的声强是大量声波声强叠加而成的。由于反射作用，这些声波不断衰减，并在房间的各个方向传播，叠加时声波来自不同的方向，并且具有不同的到达时间和大小。因此，房间里某点的混响稳态声级是混响部分所有声波声强的叠加，如图 6.9 所示。由于混响声的这种特点，房间的混响声部分通常称为混响声场。

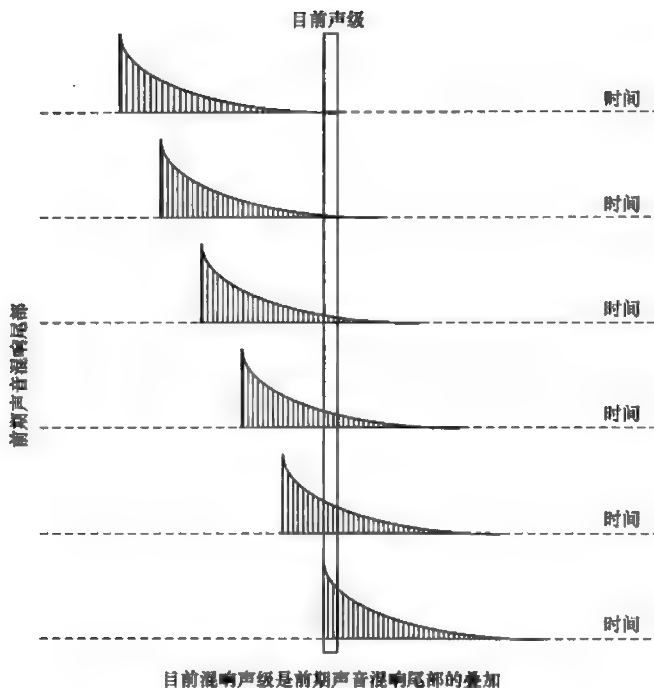
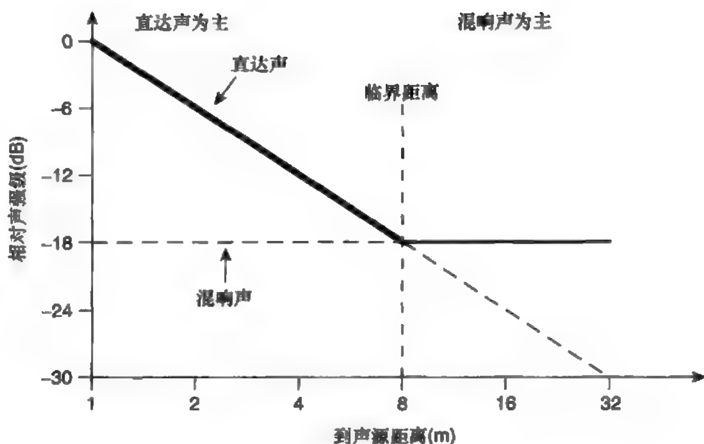


图 6.9　混响声场稳态声级的产生

6.1.6　混响声与直达声之间的平衡

混响声场的这种表现会产生两个结果。首先，直达声与混响声之间的平衡随听音者相对声源的位置而变化。这是因为混响声压级不随听音者位置变化，而直达声声压级却随着听音者与声源之间距离的变化而变化，这些现象如图 6.10 所示。图中表示出混响声与直达声的相对声级随到声源距离变化的特性。图中还可以看到，当到声源的距离达到某一值时，混响声开始大于直达声，这个距离就是直达声和混响声相等时的距离，称为临界距离。

图 6.10　直达声和混响声声强随距离变化特性



6.1.7　稳态混响声压级

其次，由于在稳态时，任意时刻的混响声强是所有混响声尾部声强之和，因此总声级由于这种叠加而增大。混响声压级与声波被吸收的速度快慢有很大关系。吸声越小，则声音的延续时间越长，叠加后产生的稳态声压级越大。如果房间的平均吸声系数为 a，则混响声强可由下式计算：

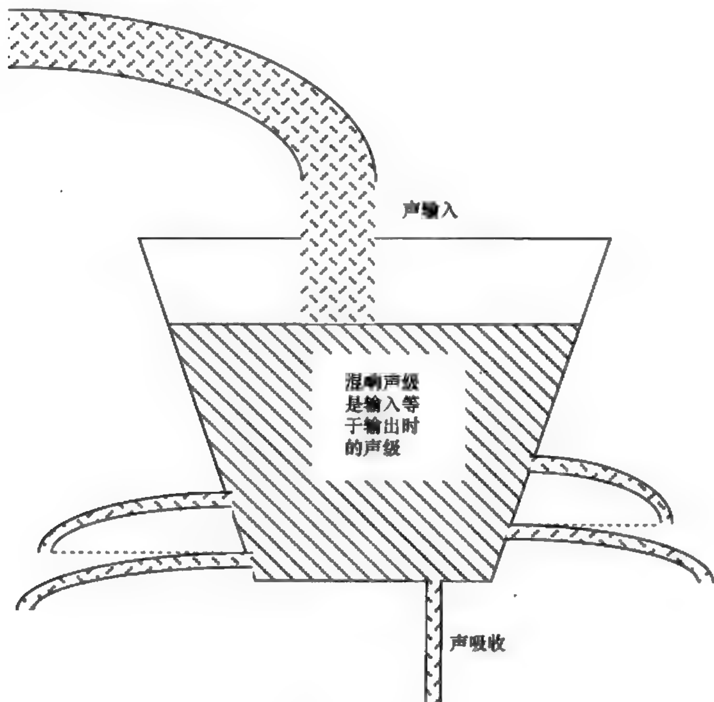
$$I_{reverberant} = W_{Source} 4\left(\frac{1-a}{Sa}\right) \tag{6.6}$$

其中，$I_{reverberant}$ 为混响声强（W/m^2），S 为房间的总表面积（m^2），W_{Source} 为声源声功率（W），a 为房间的平均吸声系数。

式（6.6）的产生是以以下实事为基础，即在平衡状态时损失的声能等于输入混响声场的能量。由于当声波撞击墙面时部分能量被吸收，因此声能被吸收的速率与表面积和平均吸声系数的乘积 Sa 成正比。这与图 6.11 所示的水倒入漏斗的情况很相似，当水流出的速率等于流入的速率时，水位达到最大值。提供给混响声场的能量也与吸声有关，因为如果吸声量越大，则直达声从表面反射回来提供给混响声场的能量就越少。请记住第一次反射之前的声音为直达声！因此，提供给混响声场的声能与第一次反射后残留的声能成正比，即正比于 $(1-a)$，其中，a 代表

第一次反射后被吸收的声能。上述两种因素综合之后得到式（6.6）中的 $(1-a)/Sa$，式（6.6）中的系数 4 与声音从全方位到达房间的墙面有关。从式（6.6）还可以看到一个有趣的现象，似乎混响声级只与总吸声面积有关，但实际上表面积和体积是有关联的，其中一个包含着另一个。事实上，由于当体积减小时房间的表面积变小，因此，尽管较小房间的混响时间总是较短，但在相同平均吸声系数的情况下，其混响声级却较高。另一种理解这种现象的方法是，认识到在较小的房间有较小的体积供声能传播，就像在一片较小的土司上涂抹一块黄油，因此，较小房间的声能密度或声级一定较高。然而，由于小房间里声波每秒撞击墙面的次数较多，其声能衰减速度要比大房间快。

图 6.11　混响声场声强级的漏斗模型



将式（6.6）中的 $(1-a)/Sa$ 取倒数后，得到一个量，称为房间常数，一般用 R 表示，即：

$$R=\frac{Sa}{(1-a)} \qquad (6.7)$$

其中，R 为房间常数（m^2），α 为房间的平均吸声系数。

将房间常数代入式（6.6）后，式（6.6）简化为：

$$I_{\text{reverberant}} = W_{\text{Source}}\left(\frac{4}{R}\right) \tag{6.8}$$

混响声强级可表示为：

$$SIL_{\text{reverberant}} = 10\lg\left(\frac{W_{\text{Source}}}{W_{\text{ref}}}\right) + 10\lg\left(\frac{4}{R}\right) \tag{6.9}$$

α 是一个取值为 0 ~ 1 的系数，较小表面积的房间与较大表面积的房间相比，其混响声级较大。但是，必须注意不要把这个结论用于某些极端情况。例如，一个窄而长的柱形房间可能有较大的表面积，但是，用式（6.6）不会预测到正确的混响声级。因为在这种情况下，声波到达各个表面的概率并不相等，因此必须对平均吸声系数进行修正，使式（6.6）的计算结果更接近实际值。因此，要注意在使用式（6.6）时存在一个前提条件，即混响声以相同的概率论从各个方向到达各个表面，这就是扩散声场的假设或条件。一般来说，扩散声场的假设是合乎实际的，可以作为房间声学设计的一个目标。但是，也有一些情况不满足扩散声场条件，例如在工作频率较低的情况下。

下面举例说明在用扬声器作为声源时，当改变吸声系数和吸声面积时对混响声场声级的影响。

例 6.4　某扬声器在 1m 处的最大声强级为 102dB。当房间总表面积为 $75m^2$、平均吸声系数为 0.9 和 0.2 时，混响声场的声级是多少？如果使面积增大 1 倍，而保持平均吸声系数不变，那么混响声场的声级变为多少？

由式（1.18）得：

$$SIL = 10\lg\left(\frac{W_{\text{Source}}}{W_{\text{ref}}}\right) - 20\lg(r) - 11\ \text{dB}$$

扬声器辐射的声功率级（SWL）为：

$$SWL = 10\lg\left(\frac{W_{\text{Source}}}{W_{\text{ref}}}\right) = SIL_{(r=1)} + 11 = 102 + 11 = 113\ \text{dB}$$

混响声场的声级为：

$$SIL_{reverberant}=10\lg\left(\frac{W_{Source}}{W_{ref}}\right)+10\lg\left(\frac{4}{R}\right)$$

在 2 种情况下房间常数为：

$$R_{(\alpha=0.9)}=\frac{S\alpha}{(1-\alpha)}=\frac{75\times0.9}{1-0.9}=675\text{m}^2$$

$$R_{(\alpha=0.2)}=\frac{S\alpha}{(1-\alpha)}=\frac{75\times0.2}{1-0.2}=18.75\text{m}^2$$

因此，混响声场的声级为：

$$SIL_{reverberant(\alpha=0.9)}=113+10\lg\frac{4}{675}=113-22.3=90.7\text{dB}$$

$$SIL_{reverberant(\alpha=0.2)}=113+10\lg\frac{4}{18.75}=113-6.7=106\text{dB}$$

面积加倍的结果是使房间常数以同样的倍数变化，因此：

$$SIL_{reverberant}=10\lg\left(\frac{W_{Source}}{W_{ref}}\right)+10\lg\left(\frac{4}{2R}\right)$$

$$=113+10\lg\left(\frac{4}{R}\right)-3$$

因此，面积加倍的结果是使混响声场的声级分别减小 3dB。

可见，混响声场的声级受到平均吸声系数的极大影响。第 1 种情况是极端“沉寂”的声学环境的典型代表，例如，一些演播室的声学环境；第 2 种情况代表起居室的平均声学特性。为了获得同样的声压级，第 1 种情况需要的扬声器声输出比第 2 种情况大约高 15dB。当音乐家在房间里演奏时，如果混响较强，他们会感受到声压级的提升。正因为如此，音乐家不喜欢在混响较弱的房间里演奏，而更愿意在一个能够帮助提高声输出的环境里演奏，这也是人们喜欢在浴室里歌唱的原因。但是，在演播室录制节目时，录音师偏好混响衰减较快的声学环境，因为此时提高节目质量是主要目的，而较强的混响可能会掩盖声压级较弱的声音细节。

6.1.8 临界距离的计算

混响声场一般是扩散声场，声波到达房间不同位置的机率相等。就像前面提到的，无论在房间的什么位置、什么时刻，我们听到的都是混响声场的总声能。正因为如此，我们可以认为声强级和声压级是相等的，即：

$$SPL_{\text{reverberant}} \approx SIL_{\text{reverberant}} = 10\lg\left(\frac{W_{\text{Source}}}{W_{\text{ref}}}\right) + 10\lg\left(\frac{4}{R}\right) \tag{6.10}$$

临界距离（critical distance）是指混响声压级和直达声声压级相等的位置，也可以通过上式计算。在临界距离处，直达声的声强和混响声的声强相等，利用式（6.1）和式（6.8）得：

$$\frac{Q\,W_{\text{Source}}}{4\pi r_c^2} = W_{\text{Source}}\left(\frac{4}{R}\right)$$

变形后得：

$$r_c^2 = \left(\frac{R}{4}\right)\frac{Q}{4\pi}$$

所以，临界距离为：

$$r_c = \sqrt{\frac{1}{16\pi}}\sqrt{RQ} = 0.141\sqrt{RC} \tag{6.11}$$

式（6.11）表明，临界距离仅与房间常数和声源的指向性有关。由于房间常数是表面积的函数，因此临界距离随房间体积的增大而增大。大多数人都在自己的起居室听音乐，因此下面举例计算高保真扬声器在起居室的临界距离。

例 6.5 设某无指向性扬声器自由式落地放置，房间的总表面积为 75m^2，平均吸声系数为 0.2，试问临界距离是多少？如果扬声器嵌入式安装在墙面上，临界距离变为多少？

由于扬声器为无指向性，因此 Q 等于 1。前面的例题已经计算出房间常数为 18.75，将已知数代入式（6.11）得：

$$r_c = 0.141\sqrt{RQ} = 0.141\sqrt{18.75\times1} = 0.61\ \text{m}(61\text{cm})$$

这是非常短的临界距离！如果扬声器采用嵌入式安装，则因为扬声器仅向半空间辐射声能，Q 等于 2，这时临界距离为：

$$r_c = 0.141\sqrt{RQ} = 0.141\sqrt{18.75\times 2} = 0.86\ \text{m}(86\text{cm})$$

临界距离仍然很小！

由于大多数人处在距扬声器约 2m 处听音，这意味着在一般的家庭听音环境，人们听到的来自高保真音响的声音是以混响声而不是直达声为主。因此对家用重放系统而言，混响声场的质量是影响系统性能的重要方面。但是，与此同时，混响声可能会降低语言清晰度，而直达声对提高语言清晰度起重要作用。

混响声场的声级与房间的平均吸声系数有关。大多数实物如地毯、窗帘、沙发和木板等都具有随频率变化的吸声系数，这说明混响声级也会随频率变化。在某些情况下，混响声级随频率的变化很大。因此，无论聆听录音节目还是其他形式的音乐，为了获得高保真的听音效果，都要求混响声级具有合适的频率响应特性。正如前面章节所述，决定声音音色的因素之一是所听到声音的频谱，也就是说，当听到的声音以混响声为主时，或听音者处在临界距离之外时，混响声将决定声音的音色。本章后面部分将进一步讨论这个问题。

6.1.9　声源指向性对混响声的影响

另外一个影响房间混响声场频率特性的因素是声源的指向性。大多数高保真扬声器和乐器在低频时呈现无指向性，但在高频时并不是这样。混响声场的声级不仅与平均吸声系数有关，而且与声源的指向性有关。因此，声源指向性的变化会影响混响声场的频率特性，从而影响声音的音色。下面以前面的典型起居室高保真扬声器为例对此做进一步说明。

例 6.6　某高保真扬声器具有平坦的轴向直达声频率响应，但指向性 Q 值随频率从 1 ~ 25 变化，其 1m 处轴向最大声强级为 102dB。房间的表面积为 75m^2，平均吸声系数为 0.2。试问混响声场的声压级随频率变化范围是多少？

由于扬声器具有平坦的轴向频率响应，式（6.1）决定的直达声声强应保持恒定不变，即：

$$I_d = \frac{Q W_{Source}}{4\pi r^2} \quad (6.12)$$

其中，I_d 为声强（W/m^2），Q 为声源指向性因数（以球形声源为参考），W_{Source} 为声源的声功率（W），r 为到声源的距离（m），应保持不变。因此，扬声器辐射的声功率可由式（6.12）得：

$$W_{Source} = \left(\frac{4\pi}{Q}\right) I_{d(r=1)} \quad (6.13)$$

式（6.13）表明，为了获得恒定的轴向直达声频率响应，当 Q 值增大时，声源的输出声功率就要减小。而混响声场的声强为：

$$I_{reverberant} = W_{Source}\left(\frac{4}{R}\right) \quad (6.14)$$

结合式（6.13）和式（6.14）得，扬声器产生的混响声场声强为：

$$I_{reverberant} = I_{d(r=1)}\left(\frac{4\pi}{Q}\right)\left(\frac{4}{R}\right)$$

因此，混响声场的声级为：

$$SIL_{reverberant} = 10\lg\left(\frac{I_{d(r=1)}}{I_{ref}}\right) + 10\lg(4\pi) - 10\lg Q + 10\lg\left(\frac{4}{R}\right)$$

由例 6.4 可知，房间常数 R 为 18.75，因此混响声级为：

$$SIL_{reverberant} = 102 + 11 - 10\lg Q + 10\lg\left(\frac{4}{18.75}\right)$$

当 Q=1 时：

$$SIL_{reverberant(Q=1)} = 102+11-10\lg1+6.7=106.3\text{dB}$$

当 Q=25 时：

$$SIL_{reverberant(Q=1)} = 102+11-10\lg25+6.7=92.3\text{dB}$$

因此，混响声级随频率变化范围是 106.3−92.3=14dB。

可见，当轴向保持平坦的频率响应时，指向性 Q 值增大，则混响声场减弱。这种现象对主观听音的影响相当于使用频率均衡器对高 Q 值的频段进行衰减，而这种处理方法对声音质量是不利的，一般不予采用。图 6.12 所示为典型的家用高保真扬声器产生的混响频率特性曲线。请注意曲线在中频和高频部分都呈现谷值，这是因为低频扬声器单元和高频扬声器单元都在有效工作频率上限呈现较强的指向性。混响声场能量的

谷点会使相应的扬声器声音被淹没，或在混音时使这些频段的声音难以听清。混响声场在高频端的下跌使声音听起来比较暗淡。有些厂家试图通过提高这些频段的轴向频率响应来消除这种效应，但又带来了一些其他问题。对于语言重放系统而言，混响声场随 Q 值的增大而减小的现象有助于提高直达声相对混响声的声压级，从而改善语言清晰度。

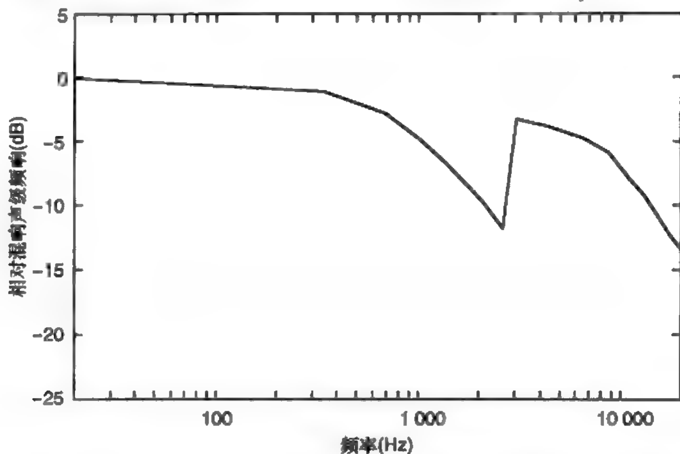


图 6.12　家用两分频高保真扬声器混响频率特性

尽管如此，许多专业录音棚的控制室还是设法利用各种声学手段来获得平坦的扬声器轴向频率响应。但对于家庭听音者而言，做到这一点却很困难。

6.1.10　混响时间

混响声场的另一特性是声场能量在一定时间内会衰减掉，这是因为每次声波与墙面撞击时都会由于吸声损失一部分能量。房间内混响声场衰减到一定程度所需的时间称为混响时间。混响时间是反映室内声场的一个重要技术指标。如果房间里的混响迅速消失，我们会感到这个房间很“沉寂”，在这样的环境里听音乐或演奏音乐对改善听闻效果没有任何益处；另一方面，如果房间混响衰减缓慢，我们感到这个房间是“活跃”的，活跃的房间比沉寂的房间更适合现场音乐的聆听或演奏。然而，当聆听的录音节目本身带有较强的混响时，声学特性沉寂的房间常常更为合适。但是，就像生活中许多事物一样，混响也必须适度，而事实上最

合适的混响时间取决于所演奏音乐的性质。例如，Scarlatti 或 Mozart 的小型复调音乐曲目适合在较小混响时间的环境演奏，而 Wagner 或 Berlioz 的大型浪漫式作品适合在较大混响时间的环境演奏。混响时间较大的情况经常出现在大教堂、溜冰场和火车站。这些场所会使大多数音乐听起来模糊不清，然而当在大教堂里聆听复调声乐，例如，Palestrina 的作品时，产生的效果会使人欣喜若狂！这是因为作曲家在创作时就已经考虑到其演奏的声学环境了。由于混响时间对房间内音乐的听音起如此重要的作用，也由于语言和不同类型的音乐对混响时间有不同的要求，人们对混响时间进行了大量的研究。事实上，当 Wallace Clement Sabine 在 1898 年发现对房间混响时间进行定量计算的方法，并以此实现对房间混响时间预测后，室内声学的研究向前迈进了一大步。在 20 世纪的前半个世纪，大多数厅堂的声学设计仅仅围绕这个参数进行，有些取得了成功，有些则非常失败。现在厅堂声学设计考虑了更多的物理声学和心理声学因素。

6.1.11 混响时间的计算和预测

混响声场消失所需的时间不仅与墙面的吸声系数有关，而且与房间内两次反射的时间间隔有关，由此我们可以推导出室内混响时间的计算公式。首先需要确定两次反射之间的时间间隔，这可以通过计算平均自由程（mean free path，缩写为 MFP）得到。平均自由程即声波处在所有可能的入射角和到达所有可能的位置（扩散声场状态）时房间内两次反射之间的平均路程。对于一个矩形房间，平均自由程可由以下公式计算：

$$MFP=\frac{4V}{S} \tag{6.15}$$

其中，MFP 为平均自由程（m），V 为房间体积（m^3），S 为房间总表面积（m^2）。

两次反射的时间间隔可由式（6.15）计算出来，即将其除以声速得：

$$\tau=\frac{4V}{Sc} \tag{6.16}$$

其中，τ 为两次反射的时间间隔（s），c 为声速（m/s）。

式（6.16）给出两次反射的时间间隔。声波每次撞击墙面时被吸收的能量比例为 α，α 即为前面提到的平均吸声系数。如果被墙面吸收的能量

比例为 a，则由墙面反射回来、即将撞击下一个墙面的能量比例为（1-a）。由于每次反射损失的能量为 a 倍，因此反射后的能量与初始能量之间的关系是按指数规律衰减。声波经过与墙面的多次反射后，利用两次反射之间的时间间隔和声能的指数衰减规律，可以推导出声能从初始状态衰减到一定比例所需时间的计算公式，详见附录 3。

衰减的比例有多种选择，但是最常用的是声能衰减 60dB 或 10^6。声能衰减 60dB 的混响时间，记为 T_{60}，可由下式计算（见附录 3）：

$$T_{60}=\frac{-0.161V}{S\ln(1-a)} \tag{6.17}$$

其中，T_{60} 为 60dB 混响时间（s）。

式（6.17）称为艾润（Norris-Eyring）混响公式，式中的负号与自然对数计算产生的负号相抵消，使混响时间的计算结果为正值。请注意当采用其他衰减比例计算混响时间时，其公式与式（6.17）的差异仅仅是所乘的系数不同。混响时间计算公式的推导是以统计学为基础的，因此式（6.17）背后存在一些重要假设，这些假设是：

（1）声波以相同的机率从各个方向到达墙面的各个位置，即声场为扩散声场。这个条件是为了产生和利用房间平均吸声系数的概念。请注意从主观听音上看，这也是所期望达到的目标，人们更愿意在扩散声场中聆听或演奏音乐。

（2）平均自由程概念成立。这也是利用平均吸声系数的条件。此外，它也意味着房间的形状不能太极端，即这种统计分析方法不适用于类似于长隧道的房间形状。然而，对于大多数实际房间，平均自由程的计算公式是适用的。

6.1.12　房间大小对混响时间的影响

从式（6.17）还可以得出当不考虑吸声量时房间大小对混响时间的影响。式（6.17）表明，混响时间是房间表面积和体积的函数，其中表面积决定了房间的总吸声量，体积决定了两次相邻反射之间的时间间隔。引入房间的线度 L，则表面积和体积随线度变化的特性如下：

$$V\propto L^3$$

$$S \propto L^2$$

而两次反射的平均时间间隔和混响时间随线度 L 的变化特性如下：

$$\frac{V}{S} \propto \frac{L^3}{L^2} \propto L$$

可见，当平均吸声系数保持不变时，混响时间随房间尺寸的增大而增大。在普通房间里，吸声往往是由地毯、窗帘和人等产生，一般可以认为其所占面积的比例不变（平均吸声系数大致相同）。因此，一般来说，大房间比小房间具有较长的混响时间，这也是我们主观评判房间大小的依据之一。另一个评判依据是早期反射声与直达声的时间间隔。我们常常看到人们更喜欢在“大”声学环境而不是“小”声学环境里听音，这里“大”、“小”其实是指混响时间的大小。现在，利用高质量的电子混响系统，也可以在小房间产生长混响，但是，在这样的环境听音总给人不自然的感觉，因为听觉感受和视觉是相矛盾的。这是因为当人们看到所在房间的外观时，凭借经验期待听到短混响的声音，而实际听到的与此不相符。显然，通过闭上眼睛可以去除视觉的影响，恢复在大房间的听音幻觉！

下面举例用式（6.17）计算混响时间。

例 6.7　某房间的表面积为 75m²，体积为 42m³，平均吸声系数分别为 0.9 和 0.2。试问混响时间分别是多少？如果房间尺寸加倍，而保持平均吸声系数不变，混响时间有什么变化？

利用式（6.17），代入已知数，当 α=0.9 时：

$$T_{60} = \frac{-0.161V}{S\ln(1-\alpha)} = \frac{-0.161 \times 42}{75\ln(1-0.9)} = 0.042\text{s}\ (42 \times 10^{-3}\text{s})$$

这个混响时间非常小！当 α=0.2 时：

$$T_{60} = \frac{-0.161V}{S\ln(1-\alpha)} = \frac{-0.161 \times 42}{75\ln(1-0.2)} = 0.43\text{s}$$

这是典型起居室的混响时间。

当房间尺寸加倍时，体积和表面积之比也加倍，因此新的混响时间增大倍数为：

$$\frac{V_{\text{doubled}}}{S_{\text{doubled}}} = L_{\text{doubled}} = 2$$

新的混响时间为原来的 2 倍，即：

当 a=0.9 时，$T_{60\ doubled}=T_{60}\times 2=0.042\times 2=0.084$s

当 a=0.2 时，$T_{60\ doubled}=T_{60}\times 2=0.43\times 2=0.86$s

6.1.13　短混响引起的问题

当房间的吸声较强时，混响时间会非常短，这就产生了一个有趣的问题。请记住上述混响时间计算公式的使用条件是声波以相同的机率撞击房间里所有的墙面。在我们的例题中，房间两次反射的平均时间间隔用式（6.16）计算得：

$$\tau=\frac{4V}{Sc}=\frac{4\times 42}{75\times 344}=6.51\ \text{ms}\ (0.006\ 51\text{s})$$

如果将例 6.7 中计算的当 a=0.9 时的混响时间与两次反射的平均间隔时间相除，则得到混响时间内的平均反射次数为：

$$N_{\text{reflections}}=\frac{T_{60}}{\tau}=\frac{42\times 10^{-3}}{6.51\times 10^{-3}}=6.45$$

这样的反射次数是根本不够的！在这种情况下，混响声场其实并不存在，房间里衰减的只是一些早期反射声，混响声场的有关概念根本不适用。为了获得混响声场，反射次数应该远大于 6 次，一般认为获得混响声场的反射次数应不小于 20 次，而精确界定反射次数的下限显然是很困难的。许多演播室和控制室是按强吸声设计的，因此很难产生混响声场。

6.1.14　简单混响时间计算公式

虽然艾润公式经常用于混响时间的计算，另外还有一个较简单的混响时间计算公式，称为赛宾（Sabine）公式，是以其推导者 Wallace Clement Sabine 的名字命名的，也经常被使用。虽然赛宾公式最初是通过考虑房间的平均声能损失、解简单的微分方程推导出来的，但它也可以从艾润混响公式推导出来，从而更清楚地看到使用赛宾公式的条件。艾润混响公式表示如下：

$$T_{60}=\frac{-0.161V}{S\ln(1-a)}$$

上式的困难在于需要计算（1-a）的自然对数。但是，自然对数可以用以下无限级数表示：

$$T_{60}=\frac{-0.161V}{S\left(-a-\frac{a^2}{2}-\frac{a^3}{3}-\cdots-\frac{a^n}{n}-\cdots-\frac{a^\infty}{\infty}\right)} \quad (6.18)$$

因为 $a<1$，所以上述级数总是收敛的。而且，当 $a<0.3$ 时，级数值与 a 值的误差小于 5.7%，这时，式（6.18）可用下式近似：

$$T_{60}=\frac{-0.161V}{S(-a)}=\frac{0.161V}{Sa} \quad (6.19)$$

式（6.19）称为赛宾混响公式，它不仅经常被使用，而且是最早出现的混响时间定量计算公式。W.C. Sabine 在大量实验测量的基础上发现了这个混响时间计算公式，从此开启了建筑声学科学研究的大门。公式（6.19）使用方便，只要满足 $a<0.3$ 的条件，就可以得到足够精确的计算结果。对于许多实际房间这样的条件一般能够得到满足。但是，随着平均吸声系数的增大，计算误差也随之增大，极端情况是 $a=1$，即没有墙面时的混响时间！

6.1.15　实际混响声场

正如前面提到的，这些公式后面的假设是，混响声场是统计上随机的扩散声场。然而，在有些情况下，声场并不满足这个条件。图 6.13 所示为理想情况下混响声场能量的分贝（dB）值随时间衰减的曲线。可见，在这种情况下，衰减曲线是一条光滑的直线，表明声场能量按指数规律衰减，且每秒衰减的分贝数相等。另一方面，图 6.14 表示两种混响声场不是扩散声场的情形。在第一种情形里，所有的吸声集中在两个表面，例如，一个地面铺着地毯、天花板上安装吸声砖，而四面墙上什么都没有的办公室，声音在两个吸声面之间迅速衰减，而在墙面之间的衰减要慢得多。第 2 种情形是由两个相邻的空间组成，例如教堂里的正厅和耳堂，或者音乐厅里大厅和挑台下面的空间。在这种情况下，声能在两个空间之间不能

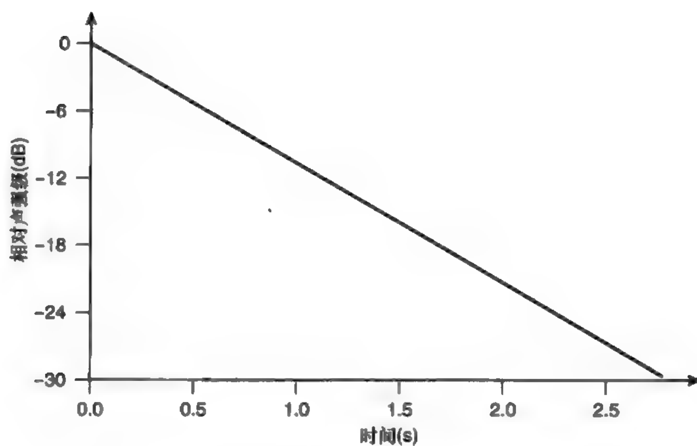


图 6.13　混响随时间衰减的理想曲线

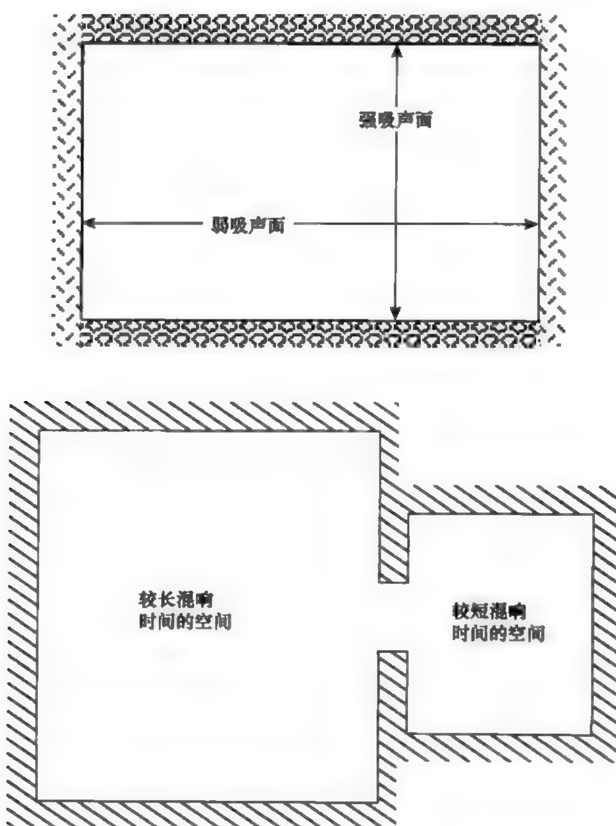


图 6.14　2种产生不理想混响衰减曲线的情形

完全耦合，因此声能以不同的速率衰减，分别取决于两个空间吸声的大小。在上述两种情形下，结果都是声能随时间以两个或更多不同的斜率衰减，如图 6.15 所示。产生这种衰减曲线的原因是，衰减速率较快的声波在衰减速率较慢的声波之前消失，因此衰减较慢的声波在最后时刻的声场中占主导地位。另一个混响衰减的主要声学缺陷发生在当存在两个相互平行的反射面时，如图 6.16 所示，会导致产生一连串短间隔的回声，从拟声角度称为颤动回声。它是由声能在两个墙面之间来回穿梭产生的，通过在两个平行墙之间拍手可以容易地检测颤动回声是否存在。在这种情况下，声能随时间的衰减曲线如图 6.17 所示，颤动回声的存在导致曲线中出现一系列峰值。请注意，这种特性往往和图 6.15 所示的双斜率特性共同存在，因为两个平行面之间的区域具有比其他扩散声场区域较弱的吸声。

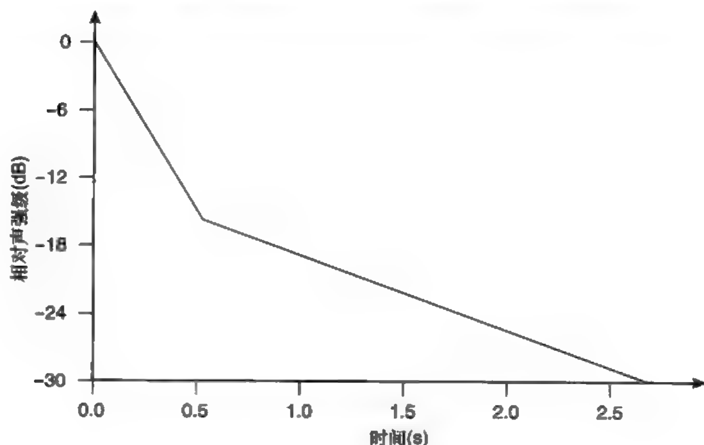


图 6.15 2个斜率的混响衰减曲线

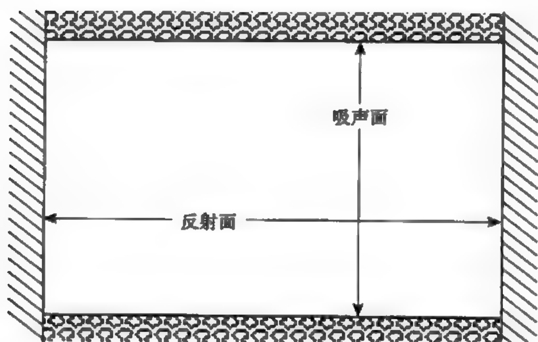


图 6.16 可能引起颤动回声的情形

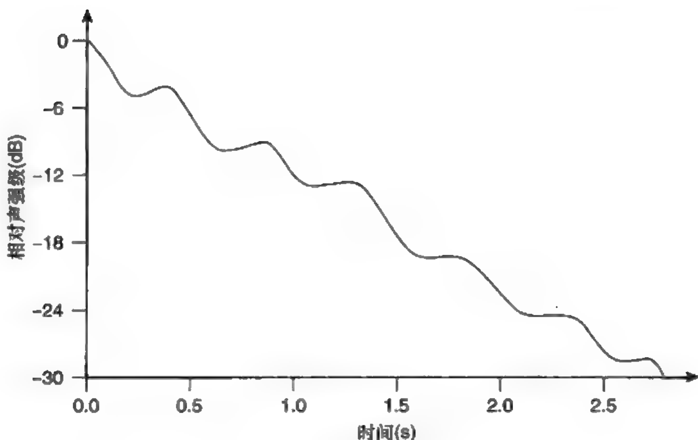


图 6.17　存在颤动回声时混响衰减曲线

6.1.16　混响时间随频率的变化

由式（6.17）和式（6.18）看出，混响时间主要与房间体积、表面积和平均吸声系数有关。然而，实际材料的吸声系数随声波频率变化，这说明当房间的体积和表面积保持不变时，房间的混响时间将随频率变化。前面介绍过，当吸声系数变化时，混响声压级也随之变化，因此会带来主观感觉上音色的变化。同时，声能衰减过程也会带来音色的变化。举一个极端的例子来说，假设某个频率的衰减速率比其他频率小得多，那么，当其他频率消失后，这个频率的声音将占主导作用，房间里就会听到这个频率的振铃声，并且由于该频率的混响声级较高，使得室内稳态声压级在该频率出现明显的峰值。

表 6.1 所示为一些常用吸声材料的吸声系数随频率变化的值，请注意它们是以倍频程间隔测量的。可能有人认为以 1/3 倍频程间隔测量吸声系数从心理声学上考虑更合适。因为倍频程间隔测量可能掩盖其中对听觉有影响的一些信息，但是，由于吸声系数随频率的变化比较缓慢，因此倍频程测量足以反映其频率特性。然而，当考虑共振结构的吸声特性时，最好采用更高的频率分辨率进行测量。吸声系数表中一般不给出 125Hz 以下的吸声系数，原因之一是，在如此低的频率进行测量难度较大；另一个原因是，当频率低于 125Hz 时，房间的其他因素对室内声学特性的影响比吸声系数更重要，这个问题将在后面讨论。

表 6.1 常用材料吸声系数

材料	频率					
	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
木板条抹灰	0.14	0.10	0.06	0.05	0.04	0.03
水泥地面铺地毯	0.02	0.06	0.14	0.37	0.60	0.65
地板（木搁栅）	0.15	0.11	0.10	0.07	0.06	0.07
刷上油漆的灰泥	0.01	0.01	0.02	0.02	0.02	0.02
墙面（1/2inch 石膏板）	0.29	0.10	0.05	0.04	0.07	0.09
窗户（浮法玻璃）	0.35	0.25	0.18	0.12	0.07	0.04
木镶板	0.30	0.25	0.20	0.17	0.15	0.10
窗帘（棉质皱褶到一半面积）	0.07	0.31	0.49	0.81	0.66	0.54
空气吸声（每 m³、气温 20℃、相对湿度 30%）	–	–	–	–	0.012	0.038

当考虑到吸声系数随频率变化的特性时，艾润混响时间计算公式应修正为：

$$T_{60}=\frac{-0.161V}{S\ln\left(1-\alpha\left(f\right)\right)}$$

其中，$\alpha(f)$ 为随频率变化的吸声系数。赛宾混响公式应改写为：

$$T_{60}=\frac{0.161V}{S\alpha\left(f\right)}$$

6.1.17 具有不同墙面的房间混响时间计算

在实际房间中，可能墙面由不同吸声材料装饰而成，并且吸声系数随频率变化，这就使问题变得复杂。因为有不同面积的吸声材料，各自有不同的吸声系数，必须找到一个能够反映它们综合作用的参数。例如，一个较大面积较小吸声系数的墙面可能比较小面积但具有较大吸声系数的墙面的影响更大。在赛宾公式中，可以较容易地将不同的吸声系数和相应的墙面面积相乘，然后将所有墙面的这个乘积相加，所得的结果赛宾称之为“等效开窗面积（equivalent open window area）”。因为赛宾假设开窗的吸声系数为 1，并且用实验证实了这个假设。赛宾混响公式（6.19）的分母即为等效开窗面积，只不过它是通过房间平均吸声系数计算的。因此，当房间里有不同吸声系数的墙面时，可以用上述方法计算出等效开窗面积，然后将其代替式（6.19）中的 $S\alpha$，得到多种吸声材料房间的

混响时间计算公式如下：

$$T_{60(\alpha<0.3)}=\frac{0.161V}{\sum_{S_i}S_i\alpha_i(f)} \tag{6.20}$$

其中，$\alpha_i(f)$ 为给定材料的吸声系数，S_i 为其面积。

对于艾润混响时间公式，由于公式里不直接使用等效开窗面积，因此公式的修正稍微复杂一些。适用于不同吸声墙面的艾润混响时间公式如下（其推导过程见附录 4）：

$$T_{60}=\frac{-0.161V}{\sum_{S_i}S_i\ln\left(1-\alpha_i(f)\right)} \tag{6.21}$$

式（6.21）称为 Millington-Sette 公式。虽然式（6.21）比赛宾公式更复杂，但是，它适合于各种不同的吸声系数，并且当平均吸声系数小于 0.3 时，可用式（6.20）等效，原因如前所述。因此，在许多情况下，赛宾公式更经常得到应用。式（6.20）更便于结合吸声系数制成表格计算混响时间，建立这样的计算表格十分容易。例 6.8 为起居室混响时间计算例子。

例 6.8 某起居室的表面积为 75m²，体积为 42m³。地面是水泥上面铺地毯，天花板是木条上面抹灰泥，面积都是 16.8m²，窗户的面积是 6m²，其余是在砖墙上抹灰泥后上油漆。忽略门的吸声影响。试问其 60dB 混响时间在不同频率分别是多少?

利用表 6.1 的数据建立一个表格，如表 6.2 所示。计算每个表面的等效开窗面积及其随频率的变化值，将每个表面的贡献相加，代入式（6.20）计算出混响时间。

表 6.2 未装修的起居室吸声和混响时间计算

表面材料	面积（m²）	频率					
		125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
天花板（木板条抹灰）	16.8	2.35	1.68	1.01	0.84	0.67	0.50
地面（水泥地面铺地毯）	16.8	0.34	1.01	2.35	6.22	10.08	10.92
墙面（刷上油漆的灰泥）	35.4	0.35	0.35	0.71	0.71	0.71	0.71
窗户（浮法玻璃）	6.0	2.10	1.50	1.08	0.72	0.42	0.24
总等效开窗面积（m²）		5.14	4.54	5.15	8.48	11.88	12.37
房间体积（m³）	42						
混响时间（s）		1.32	1.48	1.31	0.80	0.57	0.55

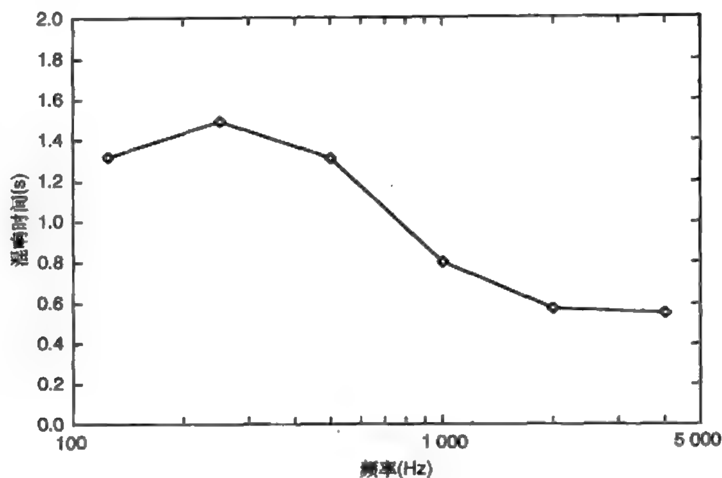


图 6.18　未装修的起居室混响时间频率特性

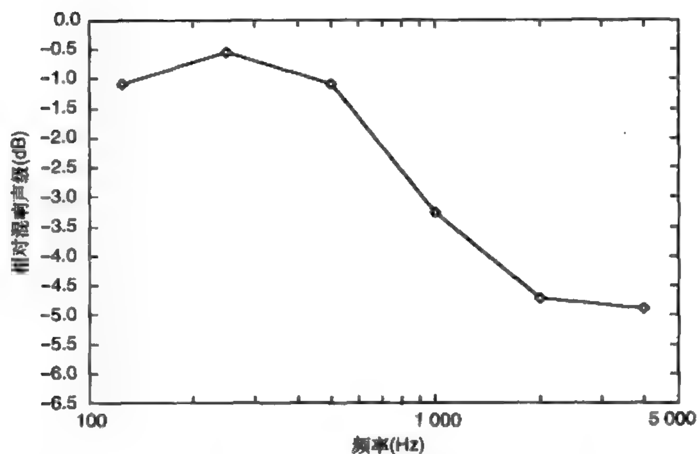


图 6.19　未装修的起居室混响声级频率特性

表 6.2 的计算结果也可用图 6.18 表示。可见，混响时间从低频的 1.49s 变化到高频的 0.55s。这是这类房间的典型混响时间特性，听起来声音有些“发毛”和“低沉”。这个房间的相对混响声级频率特性如图 6.19 所示。由图可知，低频的混响声级大约比高频高 5dB。

6.1.18　混响时间设计

例 6.8 的计算结果带来一个问题："怎样才能使混响频率响应变得平坦呢？" 答案是通过增加或去除多余的吸声材料获得所需的混响特性。在这里等效开窗面积的概念非常有用。具体方法是：由已知的房间体积和设计的混响时间计算所需的等效开窗面积，然后计算房间现有的等效开窗面积，根据其比设计值大或小，适当地去除或增加部分吸声材料。例 6.9 说明如何改善上述房间的混响频率特性。

例 6.9　在例 6.8 所述的房间中，通过增加哪种吸声材料、且增加多少能改善房间的混响时间频率特性？

可以在房间里增加低频吸声较强的材料，如木镶板。假设以 4kHz 时的等效开窗面积为设计值，则要求在整个频率范围内等效开窗面积达到约 $12.5m^2$。在上述例题中，最差的频率是 250Hz，只有 $4.5m^2$ 等效开窗面积，这意味着在该频率必须增加适量的吸声材料使等效开窗面积增加 $12.5-4.5=8m^2$。由表 6.1 可知，木镶板在 250Hz 的吸声系数为 0.25，因此所需的木镶板面积为：

$$S_{wood}=\frac{8}{0.25}=32m^2$$

表 6.3、图 6.20 和图 6.21 所示为增加上述吸声材料后对混响时间频率特性产生的影响。这时，混响时间的变化范围缩小为 0.59s ~ 0.41s，混响声级的差异也减小到 2dB。

表 6.3　**声学处理后起居室的吸声和混响时间计算**

表面材料	面积 (m^2)	频率					
		125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
天花板（木板条抹灰）	16.8	2.35	1.68	1.01	0.84	0.67	0.50
地面（水泥地面铺地毯）	16.8	0.34	1.01	2.35	6.22	10.08	10.92
墙面（刷上油漆的灰泥）	35.4	0.35	0.35	0.71	0.71	0.71	0.71
窗户（浮法玻璃）	6.0	2.10	1.50	1.08	0.72	0.42	0.24
木镶板	32.0	9.60	8.00	6.40	5.44	4.80	3.20
总等效开窗面积 (m^2)		14.74	12.54	11.55	13.92	16.68	15.57
房间体积 (m^3)	42						
混响时间 (s)		0.46	0.54	0.59	0.49	0.41	0.43

图 6.20 声学处理后起居室混响时间频率特性

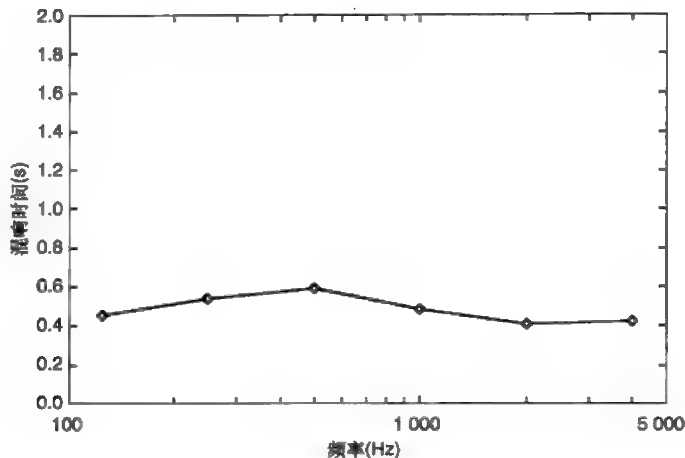
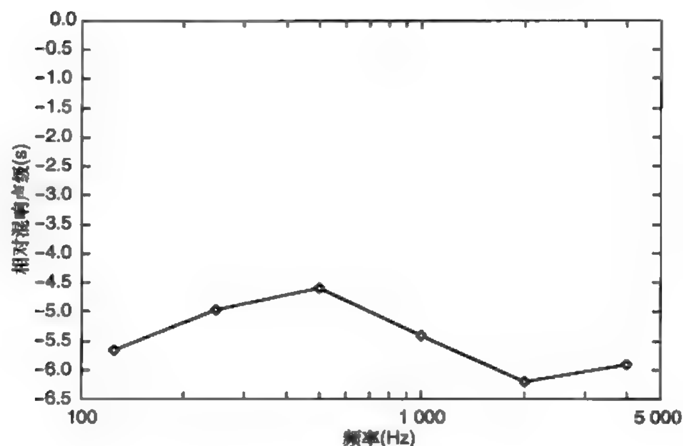


图 6.21 声学处理后起居室混响声级频率特性



然而，声学处理后所有频率的混响时间都有所下降，这是因为增加的吸声材料不仅对设定的频率起作用，而且对其他频率的吸声特性也有一定的影响。因此，在实际应用中，需要反复进行上述过程以便获得最合适的处理方案。另一点需要注意的是，上述方案仅适用于该房间，而且有时由于实施条件的限制，难以达到所设计的混响时间特性。

6.1.19 理想的混响时间特性

什么是理想的混响特性呢？我们已经看到，混响衰减时间特性应是光滑的指数衰减特性，即单位时间衰减恒定的分贝数。同时我们也知道，不同类型的音乐需要不同的混响时间。在大多数情况下，回答是“混响时间应视情况而定”，然而，也存在一些广泛适用的规则。首先，不同类型的音乐有不同的混响时间取值范围。节奏感强、起伏较快的音乐比起较慢而谐和的音乐需要更干的声学环境；其次，当演奏的空间体积增大时，所有类型音乐所需的混响时间随之增大。图 6.22 所示为理想混响时间随音乐类型和房间体积变化的特性；最后，大多数听众偏好混响时间在低频（125Hz）比中频（1kHz）提升约 40%，如图 6.23 所示。低频混响的提升有助于增加声音的“温暖感”，同时，由于低音乐器的基频往往较弱，通过增大混响声级可以提高低音乐器的音量。然而，当在演播室进行录音时，或监听录制的音乐时，低频混响提升是没有必要的，这时往往更需要平坦的混响时间频率特性。

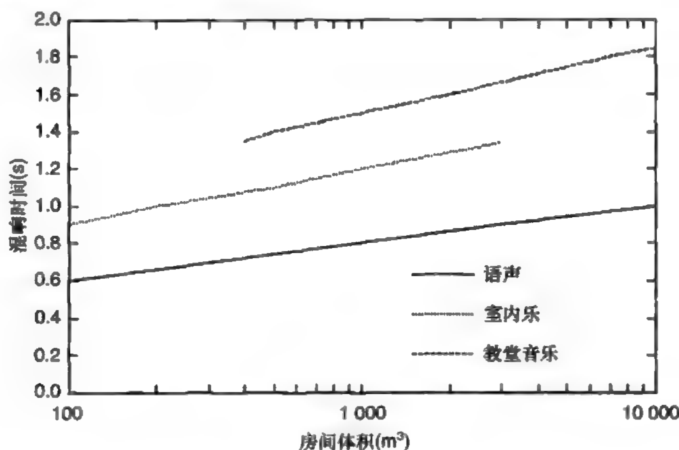


图 6.22 理想混响时间随音乐类型和房间体积变化特性

在进行室内声学设计时，关于混响还有许多其他方面问题需要考虑，但涉及太多的数学计算，因此不在此一一复述。但是，有 4 个方面问题需要在此说明，事实证明它们对室内声学设计和混响效果器设计的成败

起决定性作用。

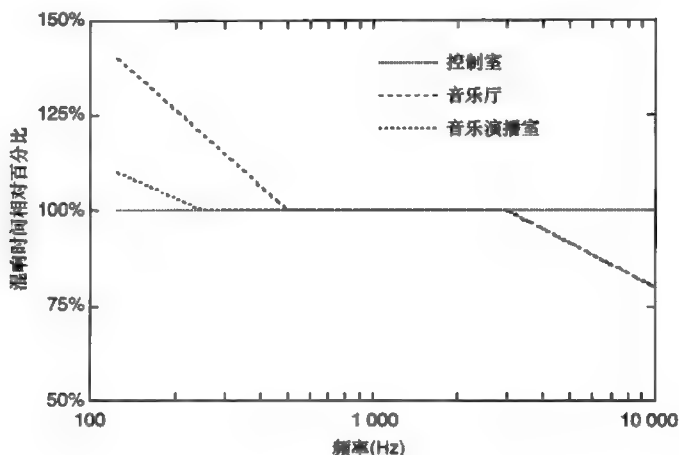


图 6.23 理想混响时间频率特性

6.1.20 早期衰减时间

第一个问题是，以衰减 60dB 测量的混响时间与听觉的关系并不是特别密切，而且在实际情况下是很难进行测定的。这是因为本底噪声的存在，使混响声级在衰减 60dB 之前就已经降低到本底噪声级，并被本底噪声所淹没。即使在维多利亚女王时代安静的小镇，那时没有交通噪声，赛宾还要用他的耳朵在深夜进行测量，以避免环境噪声对测量结果的影响。因为我们很难听到整个混响的衰减过程，因此，我们的大脑自然而然将注意力集中在所能听到的部分，所以，听觉对前 20dB ~ 30dB 的衰减曲线更加敏感。从理论上说，如果混响衰减曲线是理想的指数型，则 60dB 混响时间与早期衰减时间成正比例关系，使用 60dB 混响时间不会产生任何问题。但是，如果混响衰减曲线呈现图 6.15 所示的两个斜率的特性，则这种简单的关系就会被打破。产生的结果是，虽然 T_{60} 是描述混响声场特性的一个恰当值，由于衰减 30dB 的早期衰减速率较大，使听觉感到的混响时间要比实际混响时间短，其对听音的影响是声音听起来要比从简单测量得到的 T_{60} 值所预测的声音更“干”。因此，现代声学设计者在设计音乐厅时更关注早期衰减时间（early decay time，缩写 EDT），而

不像过去那样只关注 T_{60}。

6.1.21　侧向反射声

第 2 个对听音有较大影响的因素是来自音乐厅侧墙的密集扩散反射声，称为侧向反射声，如图 6.24 所示。侧向反射声能够使听音者产生被包围或沉浸在声场中的感觉，已被证实对改善听音者的听闻感受起重要作用。但是，侧向反射声必须是扩散的，因为镜像反射声会产生梳状滤波效应（见第 1 章），干扰声源定位，使声像产生畸变。扩散的侧向反射声是非常重要的，人们在多年前就已认识到这一点。传统上人们采用灰泥压模、壁龛和不规则的墙面装饰等获得扩散声场，近年来人们开始使用一种由许多“井”构成的扩散结构，“井”深由一系列严格定义的数值决定。但并不是只需考虑侧墙的扩散特性。传统的音乐厅通常采用鞋盒型，其体型看起来像只鞋盒，如图 6.25 所示。这种体型能为听众提供大量的侧向反射声，当配以维多利亚时期喜好的华丽石膏浮雕装饰后，能产生极好的音响效果。但是，鞋盒式音乐厅的经济效益不高，因为这种体型不能容纳某些其他体型所能容纳的那么多听众。另一种常用的体型为扇形，如图 6.26 所示。从侧向反射声角度看，它的特性和鞋盒型截然不同。这种体型能够容纳较多的听众，但是听众区不能获得足够的侧向反射声，只能在后排区域获得少量反射声，而且这些反射声太弱。通过在侧墙、顶棚安装扩散体或在空间安装悬浮扩散体，可以改善听闻效果，

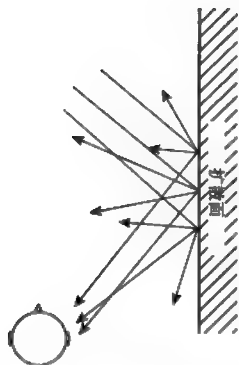


图 6.24　音乐厅的侧向反射声

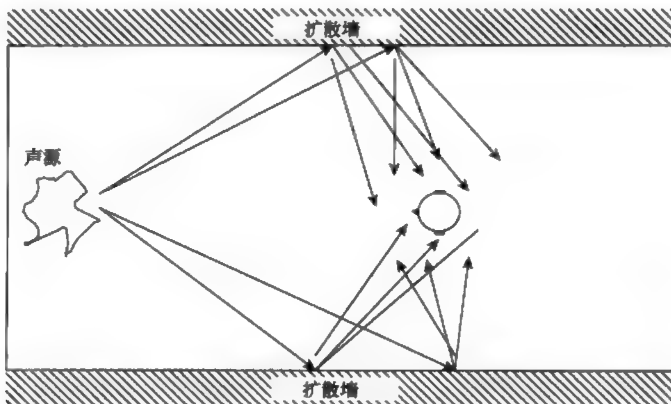


图 6.25　鞋盒型音乐厅的侧向反射声

如图 6.27 所示。然而，人们发现从听音效果上看，来自顶棚的准侧向反射声并不能代替来自侧墙的侧向反射声。因此，在处处讲经济效益的今天，获得良好听闻效果仍然是一个极具挑战性的问题。

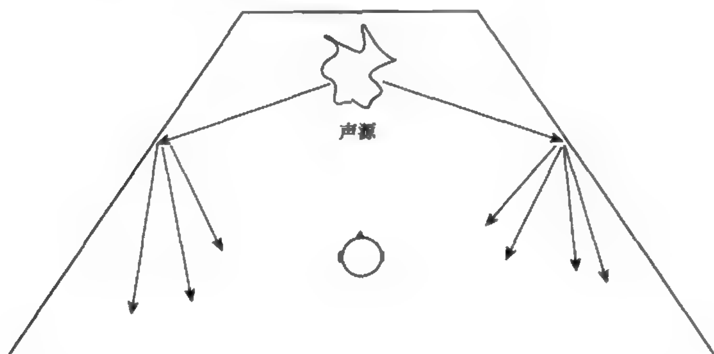


图 6.26　扇形音乐厅的侧向反射声

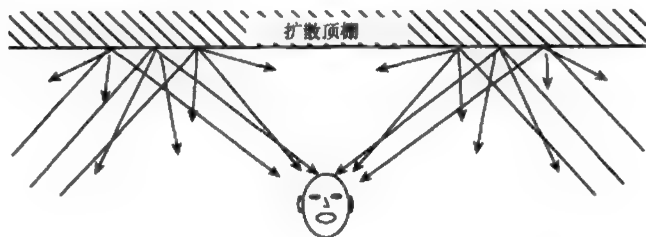


图 6.27　来自音乐厅顶棚扩散体的侧向反射声

6.1.22　早期反射声及其对演奏者的听闻支持

第 3 个需要考虑的因素，也是经常被忽略的，是演奏者的听觉感受。流行音乐演唱小组在多年前就已知道这一点，费尽心思地通过一种称为返送的技术为舞台上的每一个演员提供各自声音的平衡感。事实上，现在有些演员通过一种“耳内监听”技术直接得到返送。在许多大型特约演出中，提供给演员的返送声可以与听众听到的声音相媲美，甚至超过它们。然而，古典音乐演奏者往往只能通过厅堂获得返送，舞台上的音乐家必须依赖于来自邻近表面的反射声来听到自己和其他人的演奏声音。到达舞台的返送声音必须满足两个条件：其一，声音必须及时到达，并且声级足够大。为了满足这个条件，首先，围绕演奏者的平面必须能够将声音反射给演奏者，这一点十分重要。但是，这和提供听众最大声能

的要求产生冲突，需要相互做出妥协。最好的解决办法是充分利用向演奏者身后辐射的声音，使其经过演奏者向听众区反射，如图 6.28 所示。这样做有两个优点，一是向演奏者提供返送，二是将原本可能损失的声能导向听众区。反射给演奏者的声音最好是扩散声，这样演奏者能够听到所有乐器融合后的声音，如图 6.29 所示，从而避免镜像反射可能产生的舞台某个位置听不到某个乐器声音，而另外一些位置只能听到这个乐器声音的现象。关于声学返送的另一个重要方面是声音返回到演奏者所需的时间。理想情况是它立即返送给演奏者，有些确实能够通过地板立即返送给演奏者。经验表明，为了获得好的协同合奏效果，演员最好能在乐器发出声音后的 20ms 内听到其他演奏者的声音，这就意味着演员到周围反射面或扩散面的距离应不大于 10ms(3.44m 或 11.5ft)，允许声音在规定的时间内传播到反射面并返回。实际上当进行大型交响乐演出时，反射面到乐队的距离会比较大，但是舞台仍然有助于提供声音的返送。有时舞台空间非常大，以致反射声可能远远落后于最佳时间到达。如果到达的延迟时间大于 50ms，演员会听到回声而不去理会它们，但是，如果反射声到达的时间处在演员感觉其为声音的一部分和听到回声之间，这个反射声会对演员的听音形成严重干扰。这种"迟到的早期反射声"的副作用是使演奏者不能听到相邻其他乐器的声音，致使他们之间不能很好配合完成合奏。在某个很有声望的音乐厅，乐师们总是抱怨听不到彼此演奏的声音，不愿意在那里演奏，原因就是存在"迟到的早期反射声"。最后值得一提的是，鞋盒式音乐厅舞台往往能够提供给乐师良好的声学返送，而对于现代多功能厅，既要提供空间给布景和机械设备等，又要提供给乐师必要的返送，这是一件很难做到的事。

图 6.28　提供给演员返送的早期反射声

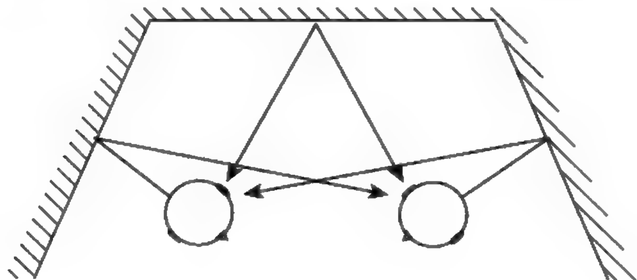
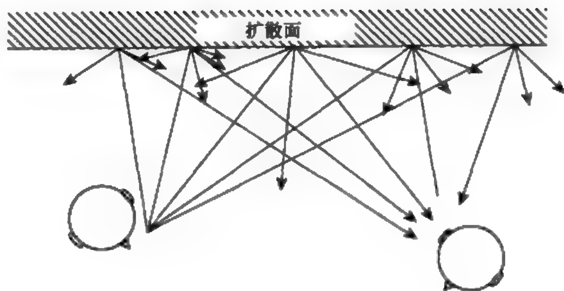


图 6.29 扩散对提供给演员返送的作用



6.1.23 空气的吸声作用

关于混响的第 4 方面问题是观察到的一种现象，它使早期的混响器设计师感到惊讶。这个现象是，在混响衰减过程中，声能不仅要经过多次反射被吸收，而且还要在空气中长距离传播，这个过程中也要损失部分声能。实际上，声波传播的路程与混响时间成正比，即 1s 的混响时间意味着声波在此衰减过程中传播的距离为 344m。尽管低频时空气的声吸收很小，但高频时的情况就不是这样。当空气比较潮湿时，烟雾颗粒和其他杂质会吸收高频声能，从而减小声音中的高频成分。这就是当人们在远处说话时声音听起来比较暗淡的原因。从混响时间和混响声级来说，这种额外声吸收产生的结果是减小高频的混响时间和混响声级，幸运的是这种影响只在频率高于 2kHz 时比较明显。由于空气声吸收与湿度、所在地烟雾的大小有关，所以有观众时高频混响时间和混响声级将不断变化，因为人们会呼出水蒸气并且出汗。请注意这是叠加在观众服装声吸收产生的静态特性之上的动态特性。显然，动态变化的大小与观众的身体状况以及通风系统的运行状况有关！由于空气的声吸收只与声波传播的距离有关，而与墙面的吸声特性无关，因此难以将其作用结合到前面讨论的混响时间计算公式中。一种较为有效的估算方法是通过计算体积和空气吸声系数的乘积将空气声吸收变换为等效吸声面积。这样做有一定的合理性。因为房间越大，则声波传播的空间越大，声波传播的路程就越长，空气声吸收就越大。空气吸声系数见表 6.1 的最后一行。可见，对于小房间，空气的声吸收是可以忽略的。当房间体积小于 $40m^3$ 时，等效吸声面积小于 $1m^2$。尽管如此，在设计人工混响器时必

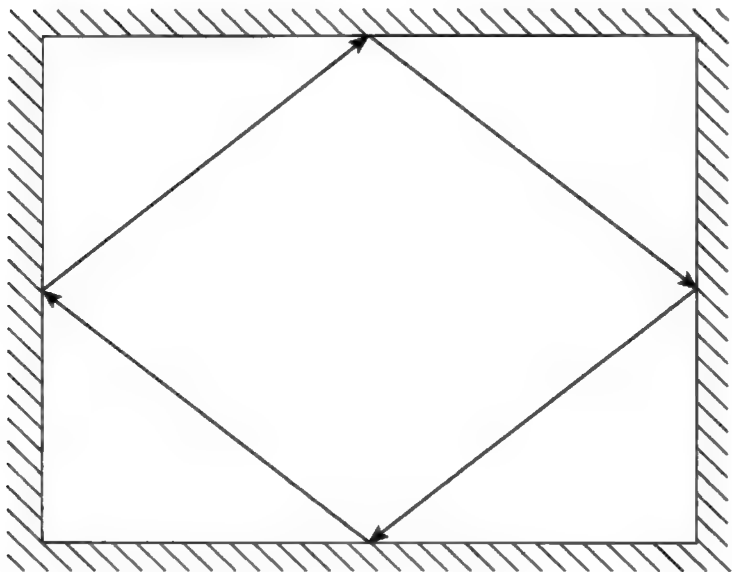
须考虑到空气声吸收的影响，否则产生的混响声可能过于明亮，听起来很不自然。

本节主要介绍了混响时间和混响声场的概念。混响时间计算公式使用条件是声场为扩散声场，如果不满足这个条件，计算结果会产生较大误差。在中频和高频时，通过无意或有意的室内声学装饰设计，一般能够满足扩散声场的条件，但是，在低频时，由于房间边界产生的驻波效应，一般很难满足扩散声场条件。

6.2　房间简正模式和驻波

当房间被一个脉冲声（瞬态声）激发后，声能通过壁面反射回来，每次反射都有一部分能量被壁面吸收，因此声能按指数规律衰减。理想状态下，声波从每个面反射的机率相等，房间内形成一个扩散声场，声能按单一的指数规律衰减，衰减常数与房间的平均吸声系数成正比。但实际情况并不总是这样，有时声能会沿着某个固定的、周期性的轨迹反射，如图 6.30 所示。如果声程正好是波长的一半，则会在房间里形成驻波。

图 6.30　房间里周期性反射路径



这些驻波（共振模式）的声压和振速在空间的分布是静态的，与房间里的其他声波有以下不同之处：

（1）驻波并不以相同的机率到达各个墙面，而是只在少数相关的墙面之间来回反射；

（2）驻波并不随机地从各个不同角度撞击墙面，而是以一定的角度入射到墙面；

（3）驻波会沿着周期性的路径回到原来的墙面，因此与频率的关系非常密切，即驻波的频率往往是离散分布的，与房间的尺寸有关。

房间驻波的另一个名称是共振模式，其发生的频率称为简正频率。由于驻波模式在空间发布是静态的，因此空间各处声压变化较大，这是所不希望存在的现象。驻波模式分为 3 种基本类型，分别在 6.2.1 ~ 6.2.3 节介绍。

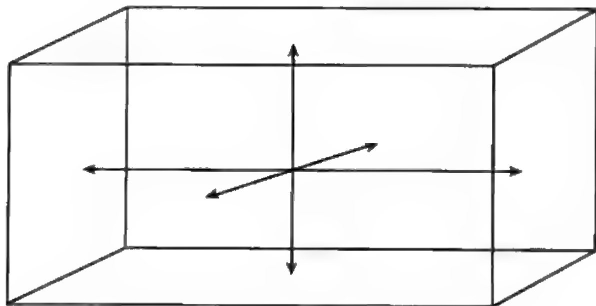
6.2.1 轴向模式

这些驻波模式发生在两个相对的墙面之间，如图 6.31 所示。其频率与房间尺寸有关。轴向模式频率由以下公式计算：

$$f_x = \frac{c}{2}\left(\frac{x}{L}\right)$$

其中，f_x 为轴向模式频率（Hz），x 为两个平面之间的半波长个数（1，2，…，∞），L 为两个相对反射面之间的距离（m），c 为声速（m/s）。

图 6.31 轴向模式传播路径



公式说明轴向模式频率有无数多个，其半波长的整数倍正好等于两个墙面之间的距离，最低模式频率的半个波长正好等于两个反射面的间距。

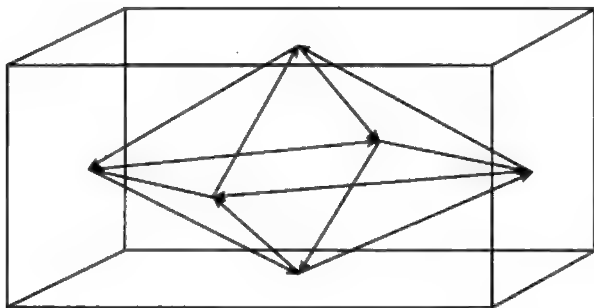
6.2.2　切向模式

这些驻波模式发生在 4 个面之间，如图 6.32 所示。其频率与房间两个方向的尺寸有关。切向模式频率可由以下公式计算：

$$f_{xy}=\frac{c}{2}\sqrt{\left(\frac{x}{L}\right)^2+\left(\frac{y}{W}\right)^2}$$

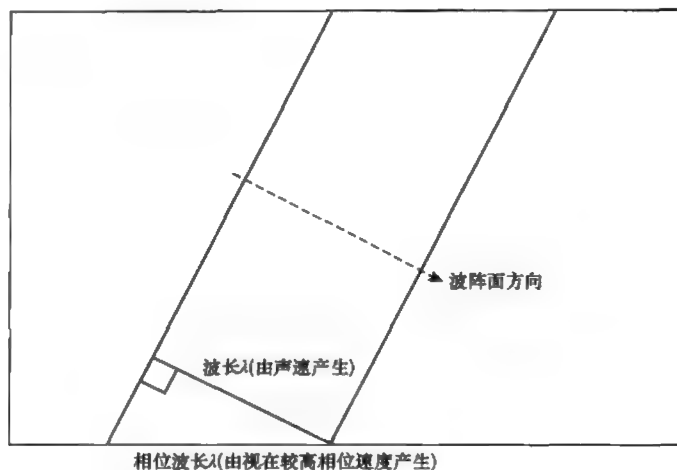
其中，f_{xy} 为切向模式频率(Hz)，x 为一对墙面之间半波长个数(1，2，…，∞)，y 为另一对墙面之间半波长个数（ 1，2，…，∞ ），L、W 分别为两对反射面之间的距离（ m ）。

图 6.32　切向模式传播路径



切向模式也有无数多个，其频率应满足的条件是，在两对反射面之间正好能够容纳整数倍半波长，结果导致最低切向模式频率比最低轴向模式的频率高，尽管从表面上看其路径更长。这是因为驻波必须与两个相对墙面的间距相适配，两个墙的间距是三角形的直角边，而不是斜边，当声波沿着斜边传播时，在房间的边界方向产生的有效波长或相位速度更大，如图 6.33 所示。因此，最低切向模式频率必须满足的条件是相位速度对应的半个波长正好等于两个墙面的间距。

图 6.33 切向模式的相位速度



6.2.3 斜向模式

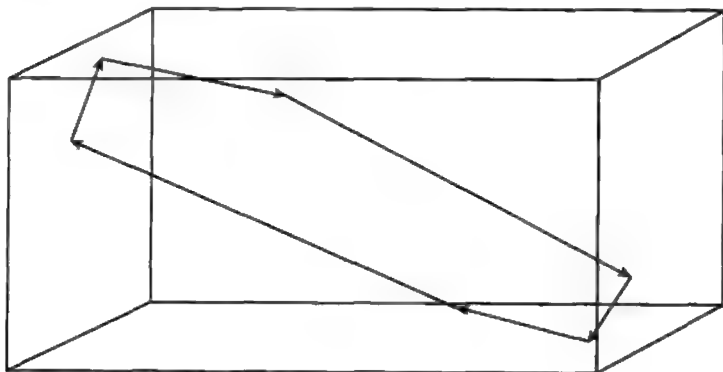
这些驻波模式发生在所有 6 个面之间，如图 6.34 所示。其频率与房间的 3 个尺寸都有关。斜向模式频率由以下公式计算：

$$f_{xyz}=\frac{c}{2}\sqrt{\left(\frac{x}{L}\right)^2+\left(\frac{y}{W}\right)^2+\left(\frac{z}{H}\right)^2}$$

其中，f_{xyz} 为斜向模式频率（Hz），x、y、z 分别为三对墙面之间半波长个数（1，2，…，∞），L、W、H 分别为 3 对反射面之间的距离（m）。

最低斜向模式频率也比最低轴向模式频率高，原因如前所述。

图 6.34 斜向模式传播路径



6.2.4　简正频率计算的通用公式

以上介绍的 3 种驻波模式的总和构成了房间里可能存在的一系列密集的驻波频率。令斜向模式中的 x、y 和 z 的取值为 0，1，2，…，∞，得到房间中所有可能的驻波模式频率计算公式如下：

$$f_{xyz}=\frac{c}{2}\sqrt{\left(\frac{x}{L}\right)^2+\left(\frac{y}{W}\right)^2+\left(\frac{z}{H}\right)^2} \tag{6.22}$$

其中，x、y、z 分别为三对墙面之间半波长个数（0，1，2，…，∞）。

上式表明，如果某个墙面间距和其他墙面间距成整数倍关系，就会出现不同模式频率相同的情况，这可能引起音质问题。因此，为了使驻波频率均匀分布，房间的尺寸最好采用非整数比关系。关于最佳房间尺寸比例的研究工作进行了许多，表 6.4 所示为一些可选用的最佳尺寸比例，但这些并不是理想房间尺寸比例的唯一选择。同时还要认识到，房间驻波模式是任何结构的包含声源的封闭空间固有存在的声波形式，不能通过改变房间形状，例如，通过使某个墙面倾斜一定角度将这些共振去除。房间形状的改变只能使共振频率值的计算从较容易变得相当复杂。Walker（1996）介绍了一种更通用的使房间尺寸比例最佳化的方法。

表 6.4　一些房间最佳尺寸比例

	高	宽	长
A	1.00	1.14	1.39
B	1.00	1.28	1.54
C	1.00	1.60	2.33

6.2.5　Bonello 准则

一定频带内的共振频率的个数随频率的增大而增大，事实上可以证明该数值与频率的平方成正比。在较大的、声学性能良好的房间里，模式密度随频率的变化是光滑的，这是评估房间共振模式形态好坏的 Bonello 准则的理论基础。这个准则试图表明，从音质主观评价上看房间共振模式的表现形态是多么重要。该准则指出，将声音频率按 1/3 倍频程划分，带宽近似于听觉临界频带，然后计算每个频带内的模式个数。

如果每 1/3 倍频程带宽的模式频率个数随频率单调上升，那么，尽管房间里存在共振，很可能仍然能够听到光滑的频率响应；如果每 1/3 倍频程内的共振频率个数随频率的升高而下降，那么房间频率响应上就可能存在听觉能够察觉的频响峰值。Bonello 准则进一步指出，模式频率重合也可能引起听觉可察觉的频率响应峰值，因此应避免在 1/3 倍频程带宽内出现模式频率重合现象，除非在此频带内同时存在至少 3 个其他非重合模式频率去平衡这两个重合模式。下面以典型起居室为例说明房间模式频率的计算方法。

例 6.10 某房间尺寸为 3.5m×5m×2.5m，试计算其最低模式频率。如果某切向模式在 3.5m 方向具有一个半波长、在 5m 方向有三个半波长，试问这个切向模式的频率是多少？斜向模式（2 2 2）的频率是多少？第一个模式重合频率是多少？

利用式（6.22）计算模式频率。最低模式频率是第一个沿着房间最长尺寸方向的轴向模式频率，即本例题中的（0 1 0）模式，其频率为：

$$f_{010}=\frac{c}{2}\sqrt{\left(\frac{0}{3.5}\right)^2+\left(\frac{1}{5}\right)^2+\left(\frac{0}{2.5}\right)^2}=\frac{344}{2}\sqrt{\left(\frac{1}{5}\right)^2}=34.4\text{Hz}$$

在 3.5m 方向有一个半波长、在 5m 方向有三个半波长的切向模式是（1 3 0），其频率为：

$$f_{130}=\frac{c}{2}\sqrt{\left(\frac{1}{3.5}\right)^2+\left(\frac{3}{5}\right)^2+\left(\frac{0}{2.5}\right)^2}=172\sqrt{0.082+0.36}=114.4\text{Hz}$$

斜向模式（2 2 2）的频率为：

$$f_{222}=\frac{c}{2}\sqrt{\left(\frac{2}{3.5}\right)^2+\left(\frac{2}{5}\right)^2+\left(\frac{2}{2.5}\right)^2}=172\sqrt{0.327+0.16+0.64}=182.6\text{Hz}$$

长度 5m 是长度 2.5m 的 2 倍，所以沿着 5m 方向的第二个轴向模式与沿着 2.5m 方向的第一个轴向模式频率重合，即：

$$f_{020}=f_{001}$$

$$f_{020}=\frac{c}{2}\sqrt{\left(\frac{0}{3.5}\right)^2+\left(\frac{2}{5}\right)^2+\left(\frac{0}{2.5}\right)^2}=172\sqrt{\left(\frac{2}{5}\right)^2}=68.8\text{Hz}$$

$$f_{001}=\frac{c}{2}\sqrt{\left(\frac{0}{3.5}\right)^2+\left(\frac{0}{5}\right)^2+\left(\frac{1}{2.5}\right)^2}=172\sqrt{\left(\frac{1}{2.5}\right)^2}=68.8\text{Hz}$$

6.2.6 共振模式的表现

从前面讨论可知，共振模式的表现不同于扩散声场的表现，存在以下一些结论：

（1）由于非随机入射，驻波模式撞击的墙面数目较少，吸声系数也比随机入射时小，因此驻波模式的声吸收并不像其他声波那样强；

（2）吸声量的减少与频率关系极大。在驻波发生的频率，声吸收较小，声音的衰减速度较慢；

（3）房间声能的衰减不再是单一的指数规律衰减形态（时间常数正比于平均吸声系数），而是存在几个衰减时间。其中最短的衰减时间一般由扩散声场产生，较长的衰减时间往往由房间共振模式产生，结果使得这些频率的声音成分过多，使房间的音质下降。

模式的能量如何随时间衰减？它与混响有什么关系？模式的声吸收对频率响应有何影响？

6.2.7 轴向模式的衰减时间

模式声能的衰减在很多方面与附录 3 中分析的声能衰减相同，主要区别在于模式的吸声系数较小。这是因为模式声波并不处于随机入射状态，吸声系数只与一些特定的墙面有关，而不是整个房间的平均值。此外，两次反射的时间间隔主要与模式的路径有关，而与房间的平均自由程无直接关系。因此，模式的衰减时间极可能与混响声的衰减时间不同。

例如，轴向模式的路径长度是由支持该模式的两个反射面间距决定的，是房间的长、宽、高尺寸之一。因此，对轴向模式而言，衰减一定时间后的声能由附录 3 中的式（A3.5）修正后计算，即将平均自由程用反射面间距代替后得：

$$1\text{s 以后的模式声能} = \text{模式初始声能} \times \left(1-\alpha_{\text{mode}}\right)^{t(c/L_{\text{mode}})}$$

其中，L_{mode} 为模式的两个反射面间距（m），α_{mode} 为模式每次反射的吸声系数。

由上式可以导出衰减 60dB 所需时间的计算公式，此公式与式（6.17）类似，即：

$$T_{60(Modal)}=\left(\frac{L_{mode}}{c}\right)\frac{\ln\left(10^{-6}\right)}{\ln\left(1-\alpha_{mode}\right)}=\left(\frac{L_{mode}}{\ln\left(1-\alpha_{mode}\right)}\right)\frac{\left(-13.82\right)}{344}=\frac{-0.04L_{mode}}{\ln\left(1-\alpha_{mode}\right)}$$

其中，$T_{60(Modal)}$ 为模式的 60dB 衰减时间（s）。

上式与式（6.17）相似，区别主要是由两个反射面间距与平均自由程不同引起的。如果模式结构的长度比平均自由程长，当吸声系数相同时，模式的衰减时间将比混响声的衰减时间长；反过来，如果模式的结构长度比平均自由程短，那么，模式的衰减时间就比混响声的衰减时间短。两次反射之间的路程长度不仅与反射面结构有关，而且与模式的类型即轴向模式、切向模式和斜向模式有关。对于轴向模式，模式路径长度 L_{mode} 为相应的房间尺寸。

6.2.8　其他类型模式的衰减时间

对其他类型模式来说，情况就更加复杂。图 6.35 所示为切向模式的传播路径。可以证明，这类切向模式的路径长度等于相关的 4 个面围成的矩形的一半长度和一半宽度构成的直角三角形的斜边，即该模式路径为：

$$L_{mode}=\sqrt{\left(\frac{L}{2}\right)^2+\left(\frac{W}{2}\right)^2}=\frac{1}{2}\sqrt{\left(L^2+W^2\right)}$$

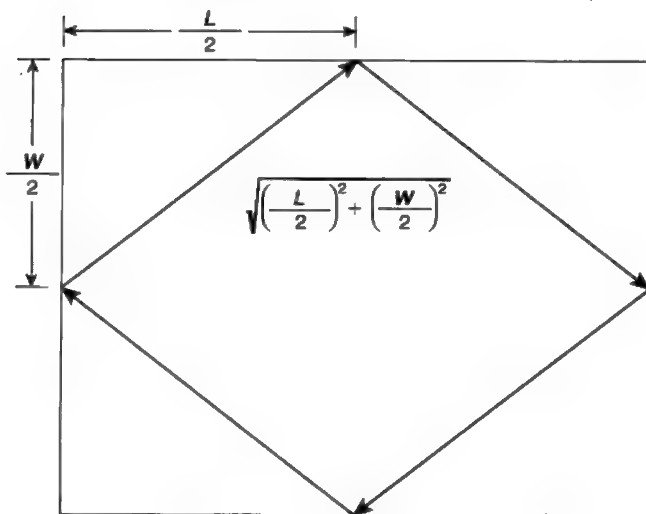


图 6.35　切向模式的路径长度

上式表明切向模式的路径长度实际上是支持该模式的 4 个面构成的矩形对角线的一半，减半是因为声波沿着该路径要经过两次反射。如此推导出来的切向模式路径可用于公式中计算模式的衰减时间。同理可以得到与 6 个面有关的斜向模式的路径长度。由于支持斜向模式的平面是 6 个，构成了一个立方体，因此必须计算两个相对的顶角之间的对角线长度。此外，由于声波沿着该路径要经过 3 次反射，斜向模式的路径长度为：

$$L_{mode}=\sqrt{\left(\frac{L}{3}\right)^2+\left(\frac{W}{3}\right)^2+\left(\frac{H}{3}\right)^2}=\frac{1}{3}\sqrt{\left(L^2+W^2+H^2\right)}$$

吸声是导致模式衰减的重要原因，它减少模式所包含的声波能量，就像声吸收对混响声场衰减的影响一样。驻波模式会引起一个有限的共振峰带宽，其带宽与吸声量成正比，如图 6.36 所示。吸声也会减小驻波中最大声压和最小声压的差值，从而减小声压在空间分布的不均匀度，如图 6.37 所示。模式的带宽可以用 60dB 衰减时间计算如下：

$$BW_{mode}=\frac{2.2}{T_{60(Modal)}}$$

其中，BW_{mode} 为模式的 −3dB 带宽（Hz）。

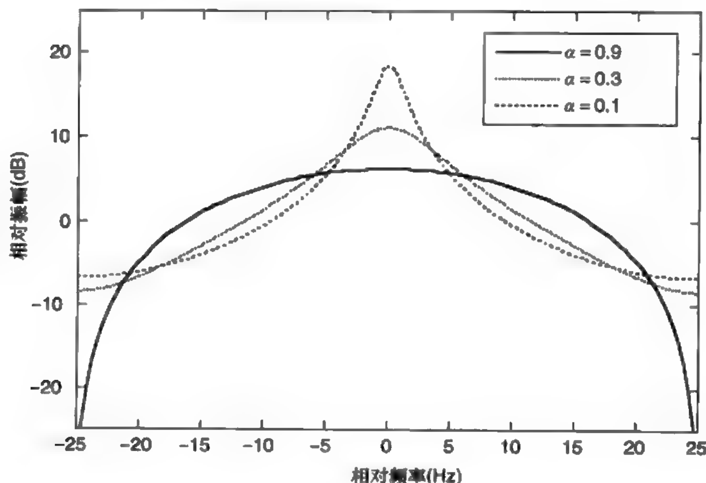
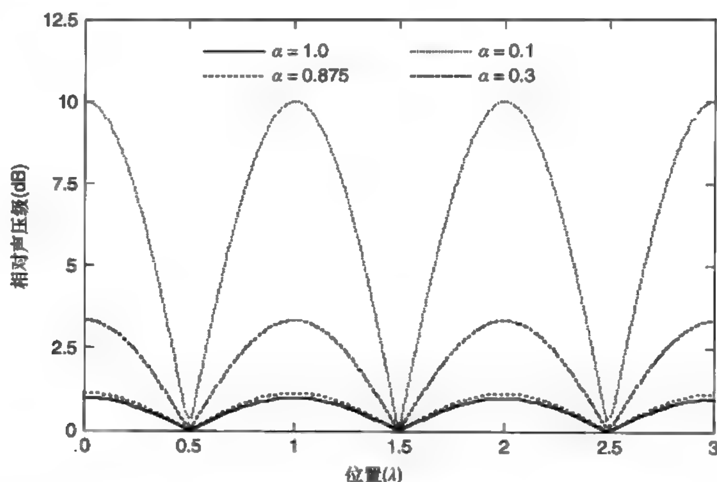


图 6.36　给定吸声系数的模式带宽

图 6.37　给定吸声系数的模式振幅空间分布

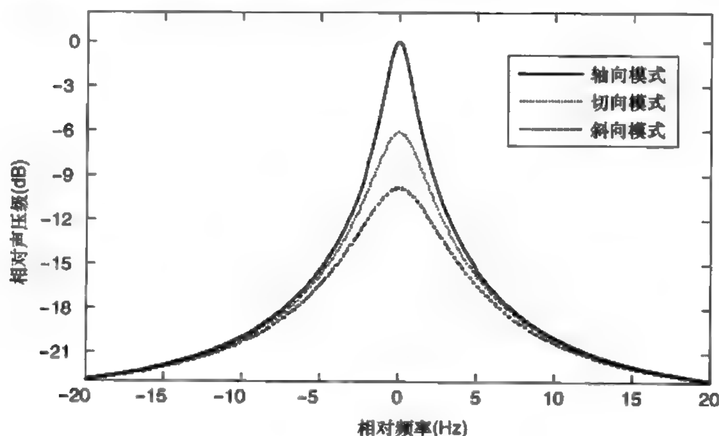


由于计算模式衰减时间比较困难，上式可以勉强用房间混响时间代替得：

$$BW_{mode} \approx \frac{2.2}{T_{60}}$$

做这样的替代比较勉强，因为混响时间是在假设声场为扩散声场的前提下计算得到的，而模式并不是扩散声场。通常模式带宽和声强级与支持模式所需的反射次数之间成比例关系，也就是说，轴向模式往往具有最强声压级，切向模式次之，最后是斜向模式，如图 6.38 所示。但是，情况并不总是这样，有时 4 个面之间的切向模式的振幅要比另外两个吸声较大的面之间的轴向模式的振幅大。

图 6.38　吸声均匀时不同类型模式的典型振幅和带宽差异



例 6.11　试计算混响时间为 0.44s 的房间的模式近似带宽。如果沿着房间 5m 长度方向的两个相对面的吸声系数等于房间的平均吸声系数 0.2，该方向轴向模式的带宽是多少？

模式的近似带宽可以计算如下：

$$BW_{mode} \approx \frac{2.2}{T_{60}} = \frac{2.2}{0.44} = 5\,\text{Hz}$$

为了回答第 2 个问题，首先需要计算模式衰减时间，即：

$$T_{60(\text{Modal})} = \frac{-0.04\,L_{mode}}{\ln\left(1-\alpha_{mode}\right)} = \frac{-0.04\times 5}{\ln\left(1-0.2\right)} = 0.9\,\text{s}$$

因此，实际该模式带宽为：

$$BW_{mode} = \frac{2.2}{T_{60(\text{Modal})}} = \frac{2.2}{0.9} = 2.4\,\text{Hz}$$

由于扩散声场的条件不再存在，因此计算模式带宽时要特别小心。在上述例题中，尽管假设吸声量在房间的分布是均匀的，模式的衰减时间和带宽与扩散声场条件下的计算结果存在较大差异，主要是因为模式的传播路径大于房间的平均自由程。在实际应用中，房间墙面各处的吸声系数很可能不同，因此估算模式的衰减时间和带宽更加困难。需要指出，当吸声不随频率变化时，模式的带宽也与频率无关，因为模式带宽只与模式衰减大小有关。

6.2.9　临界频率

由于所有房间在低频都存在驻波模式，因此总存在一个频率值，当频率低于此值时，房间里驻波模式起主导作用，声场不再是扩散声场。即使是消声室也存在有效工作频率下限。房间驻波模式对声场的影响之一是，通过对混响声场的作用使房间频率响应起伏变化。由于驻波的声压随位置变化，受驻波模式影响的房间频率响应也将随位置变化。其产生的重要结果是，在驻波起主导作用的频率范围，声场不再是扩散声场，混响时间的概念就不再适用了，而应该采用以模式衰减为基础的计算方法。但是，在什么频率会发生这种声场特性的转换呢？这个频率值能计

算出来吗？图 6.39 所示为典型的房间频率响应，图中表示出 3 种不同的频率区域。

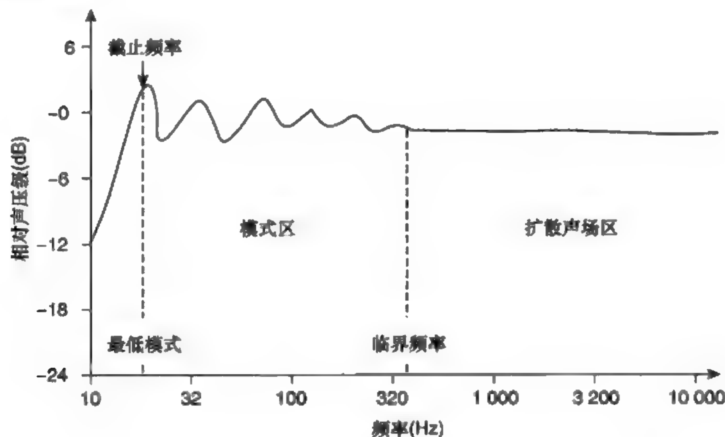


图 6.39　典型房间频率响应

（1）截止区：低于最低共振模式的频率区域有时称为截止区或压强区。在此频率区域，房间的所有尺寸都比半个波长小。这并不意味着房间里没有声波传播，但它更像一个终端没有开口的打气筒。也就是说，当房间里存在不同声源（如扬声器、乐器等）时，它产生的效果经常是（但并不总是）减小声源向空间辐射声波的能力，从而减小这些频率的声压级。这是因为声源的辐射阻抗发生了变化，使之难以向空间辐射声波。低频截止频率可简单地按下式计算：

$$f_c = \frac{c}{2L} = \frac{344}{2L}$$

其中，L 指房间的最大尺寸。

（2）模式区：紧接着截止区的是声学特性以驻波模式为主的频率区域，称为模式区。在此区域，以扩散声场为前提的分析方法肯定不适用。

（3）扩散声场区：最后一个频率区域是扩散声场存在的区域，在此区域有关混响时间的概念成立。通常在此频率范围，只要混响频率特性良好，音质将达到最佳。此时房间的驻波效应最小，听音者感受到的是一个在整个房间均匀的混响声。

从模式区变换到扩散区的分界频率称为临界频率。虽然临界频率是指某个特定的频率值，但实际上特性转换的边界是模糊的。临界频率只

是代表发生区域转换的某个特定频率。

6.2.10　声学上的“大房间”和“小房间”

临界频率（也称为“Schroeder 频率”或“大房间频率”）的概念使我们能够从声学意义上定义房间的“大”和“小”。对声学上的大房间而言，临界频率出现在房间里声源的最低频率之下；而对声学上的小房间而言，临界频率处在房间声源的频带之内。声学上的大房间的例子有音乐厅、大教堂和大型录音棚。大多数人会在声学上的小房间如卧室、浴室和起居室等听音乐或演奏音乐。由于音频工作站技术的应用和小房间的低费用，越来越多的人能够在小房间里演奏和制作音乐节目，这已形成一种发展趋势。

6.2.11　临界频率的计算

如何计算临界频率呢？主要有两种方法。第一种方法是认识到当波长达到房间的平均自由程时，驻波模式的表现可能开始增大，因为这时声波开始与所有墙面“接触”。这个方法可用于设定临界频率的下限，当频率低于此值时，如果没有采用特殊的声学处理，很难避免驻波模式效应在声场中占主导地位。基于上述方法，假设当平均自由程等于一个半波长时，驻波模式效应开始明显增大，由此得到临界频率的计算公式，即：

$$f_{\text{critical}}=\left(\frac{3}{2}\right)\frac{c}{MFP}=\left(\frac{3}{2}\right)\frac{344}{MFP}$$

其中，*MFP* 为房间的平均自由程。

上式可以方便快速地用于评估某个房间的临界频率。然而，实际临界频率可能远比计算值高。因为当房间的吸声系数较小时，可能在较高的频率仍然存在显著的驻波效应。正由于此，另一个可行的临界频率定义是以模式带宽为基础（尽管这在最初设计阶段会产生鸡和蛋的情形），同样可以产生前面介绍过的公式。原因如下所述。模式效应产生的结果是引起频率特性和空间声压分布的波动。这意味着，如果某个频率只激发起一个驻波模式，那么它引起的波动会很大；然而，如果一个频率激发起一个以上的驻波模式，那么其频率特性的波动和空间声压分布的波

动都将减小。图6.40所示为3个相邻的驻波模式叠加后的效果。由图可知，当3个相邻的模式叠加时，波动范围大大减小。用1个频率同时激发3个驻波模式的方法是增大模式带宽，直到3个模式都能覆盖到指定的频率，如图6.41所示。模式重叠效应随着频率的增大而增大，因此，将临界频率定义为当3个模式重叠时的频率，即至少3个模式被激发的频率，由下式计算：

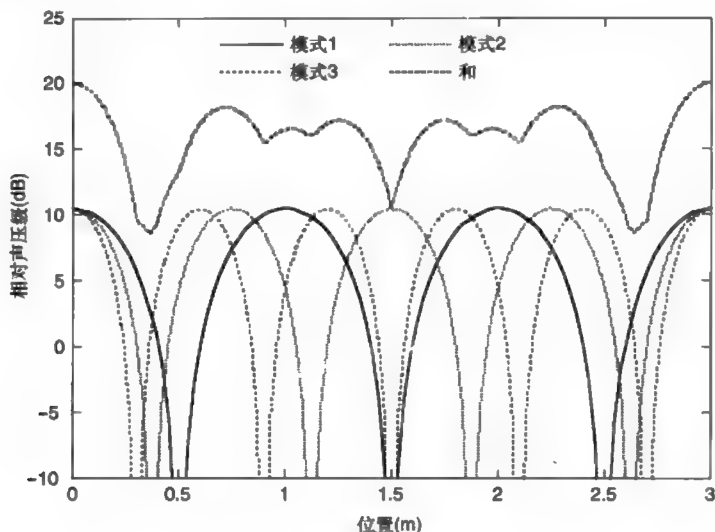


图6.40　相邻模式叠加后对声压空间分布产生的效果

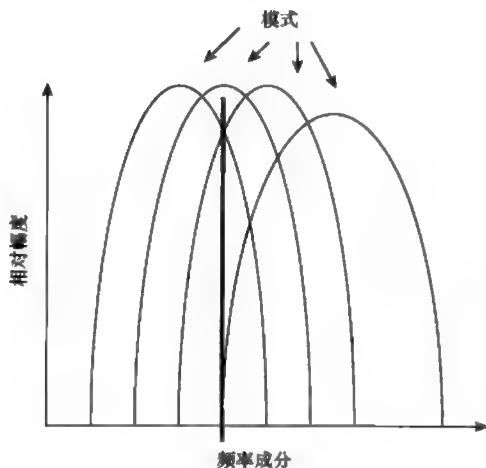


图6.41　3个模式同时覆盖某一特定频率的方式

$$f_{criticl}=2\,102\sqrt{\left(\frac{T_{60}}{V}\right)}$$

上式表明，临界频率与房间体积的平方根成反比，与混响时间的平方根成正比。而从前面章节讨论可知，当吸声系数保持不变时，混响时间与体积的立方根成正比（考虑到公式分母中面积的影响）。因此，就像我们所预测的那样，较大房间的临界频率小于较小房间的临界频率。因此，大房间从声学意义上说也是"大房间"。

下面以典型起居室为例计算房间的临界频率。

例 6.12 某房间的总面积为 75m²，体积为 42m³，试问其临界频率为多少？如果同样房间的平均吸声系数为 0.2，则临界频率变为多少？

利用第 1 个公式，代入平均自由程计算临界频率如下：

$$f_{critical}=\left(\frac{3}{2}\right)\frac{c}{MFP}=\left(\frac{3}{2}\right)\frac{c}{\frac{4V}{S}}=1.5\times\frac{344}{\frac{4\times42}{75}}=1.5\times\frac{344}{2.24}=230\text{Hz}$$

利用第 2 个公式，代入混响时间计算临界频率。首先计算混响时间如下：

$$T_{60}=\frac{-0.161V}{S\ln(1-\alpha)}=\frac{-0.161\times42}{75\times\ln(1-0.2)}=0.43\text{s}$$

因此，临界频率为：

$$f_{criticl}=2\,102\sqrt{\left(\frac{T_{60}}{V}\right)}=2\,102\sqrt{\left(\frac{0.43}{42}\right)}=213\text{Hz}$$

用第 2 个公式计算的临界频率比第 1 个公式的计算结果稍微小一些，但基本上是一致的。模式重叠的计算方法中用到了混响时间，因此也可以说存在声场是扩散声场的假设，也许在两种区域的分界频率上这种假设正好能够成立。临界频率的计算结果说明，对于这个房间，当频率低于 213Hz 时，声场的衰减必须用模式衰减时间来衡量，而不是扩散声场的混响时间。

6.3 吸声材料

吸声材料对控制房间的声学特性非常重要。本节将简要介绍影响吸声材料性能的因素以及吸声材料对空间声学特性的作用。

吸声材料基本上存在两种类型，一种是多孔吸声材料，另一种是共振式吸声结构。由于这两种吸声材料和结构的工作原理不同，因此其吸声性能也完全不同。

6.3.1 多孔吸声材料

地毯、窗帘和其他柔软的材料都属于多孔吸声材料。它们之所以能够吸声，是因为声波的振动质点与材料表面相互作用而产生的摩擦损失。从第 1 章我们已经看到，由于空气分子在稀疏状态和压缩状态之间变化，声波的振动质点存在一定的运动速度。对于给定的声压，当频率越高时，波长越短，压强变化梯度越大，质点振动速度的峰值越大。由于声波的压强梯度随着频率的增大而增大，声波与材料表面的摩擦力也随频率增大。因此，此类材料的吸声也随频率的增大而增大。显然，与声波相互作用的材料面积越大，则摩擦力越大，吸声也就越大。这就是说，单位体积包含大量纤维的多孔吸声材料，如高密度岩棉、玻璃纤维和长毛绒地毯等，往往具有较强的吸声作用。这也说明了为什么按一定比例打褶的窗帘比展开的面料能够吸收更多的声能。图 6.42 所示为典型多孔吸声材料的吸声特性曲线。由于多孔吸声材料要与声波振动质点相互作用，因此材料与墙面之间的距离、材料的厚度等对吸声有较大影响。因为在坚硬的表面如墙壁，声波质点振速为零，而在距墙面 1/4 波长处质点振速达到最大值（见第 1 章），所以，对于某些频率，其波长的 1/4 小于材料与墙面之间的距离，或者当材料直接安装在墙面上时，其波长的 1/4 小于材料厚度，吸声材料将吸收更多声能，这种效应如图 6.43 所示。虽然从理论上说，当频率增大到 1/4 波长分界点以上时，吸声系数还会随频率产生变化，但由于当到墙面的距离一定时，质点振速本身随

波长或频率变化，因此，除了材料很薄外，实际上吸声系数随频率的变化不大。

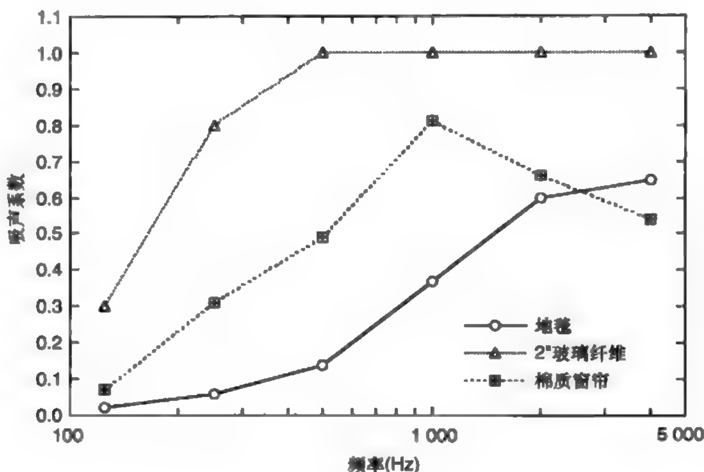


图 6.42　多孔吸声材料的典型吸声特性

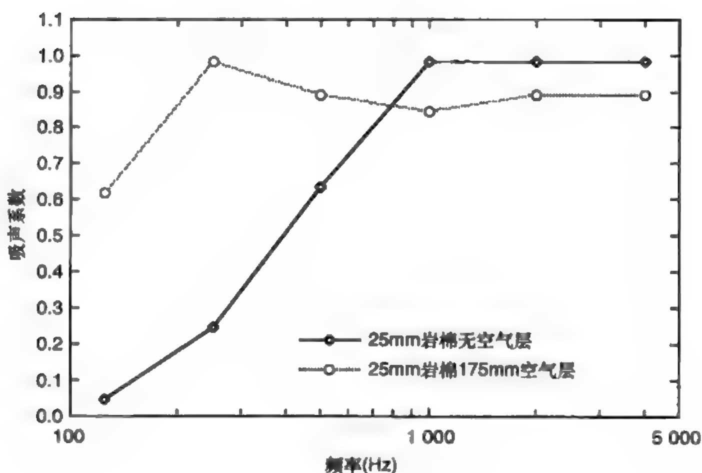


图 6.43　多孔吸声材料与墙面之间留有空气层的吸声效果

6.3.2　共振式吸声结构

共振式吸声结构有木板等，其工作原理是，声波作用使吸声体产生振动，振动能量再转换成结构内部的摩擦损耗。共振式吸声结构对声波

的声压比较敏感，因此，当将其安装在墙面上时吸声效果较好。典型的板式共振吸声结构如图 6.44 所示。对于板式吸声结构，声能是通过板内的摩擦阻力损耗掉。在后面将介绍穿孔板吸声体，它是通过孔内空气产生共振使振速增大来吸收声能的。由于吸声体是靠共振吸收声能，因此其吸声系数在低频较高，如图 6.45 所示。

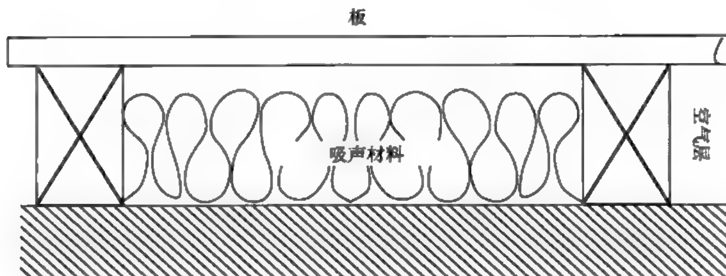


图 6.44 典型共振板吸声结构

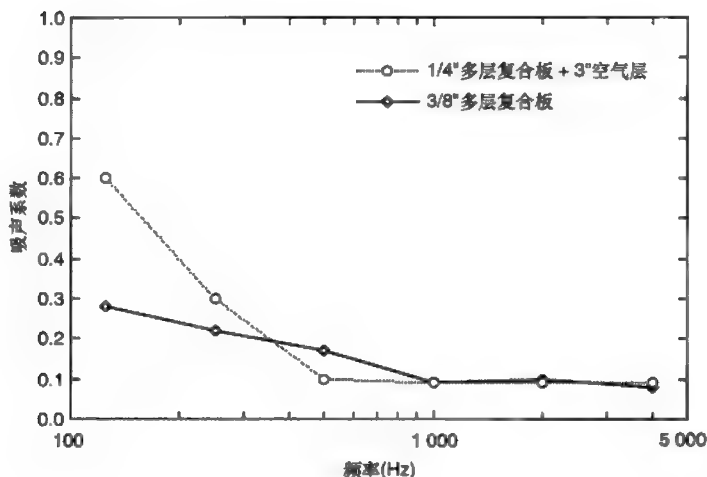


图 6.45 板式吸声结构典型吸声特性曲线

这类吸声结构的共振特性使其吸声特性能够被调整到某个低频，正好与多孔吸声材料形成互补。共振式吸声结构的最强吸声频率与板后的空气层以及面板的有效质量有关。这种吸声结构可以用一个系着质量块的弹簧系统类比，背后的空气层就像一个弹簧，其弹性系数与空气层的厚度成反比，单位面积面板的质量决定了质量块的大小。当弹性系数变小而有效质量变大时，工作频率降低。因此，较厚的空气层会使这种吸

声结构的共振频率下降。对于板式共振吸声结构，单位面积面板质量直接决定了有效质量，因此，面板越重，共振频率越低。面板吸声结构的共振频率可用下式计算：

$$f_{resonance}=\frac{60}{\sqrt{Md}}$$

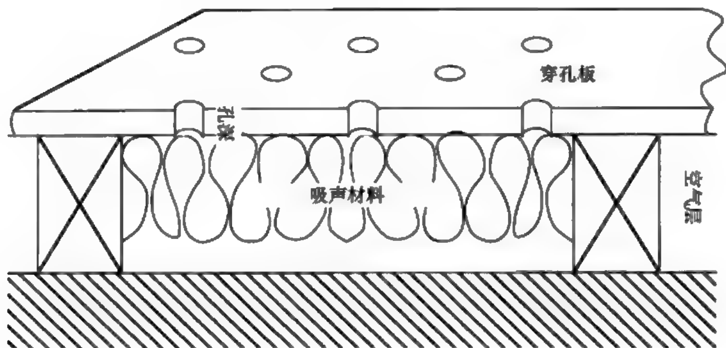
其中，M 为面板单位面积质量（kg/m^2），d 为空气层厚度（m）。

在应用上式时要注意到，它是在假设面板没有刚度的前提下成立的。对于较薄的面板，这个假设是成立的，但是当面板变厚且刚性增大后，这个公式就不太适用了。

6.3.3　赫姆霍兹共振吸声结构

另外一种共振式吸声结构是基于空腔上的空气柱发生共振的原理，就像你往啤酒瓶里吹气时发生的共振一样，这类共振吸声体称为赫姆霍兹共振吸声结构。空腔的作用相当于弹簧，空腔上的空气柱相当于质量。这种吸声结构可以由穿孔板距墙面一定距离安装构成，如图 6.46 所示。

图 6.46　典型赫姆霍兹共振吸声结构



对穿孔板而言，其有效质量与孔深和穿孔率（穿孔面积占总面积的百分比）有关。当孔深增加和穿孔率减小时，有效质量增大。图 6.47 所示为这种吸声结构的典型吸声特性曲线。这种吸声结构常用于增加中低频的声吸收。

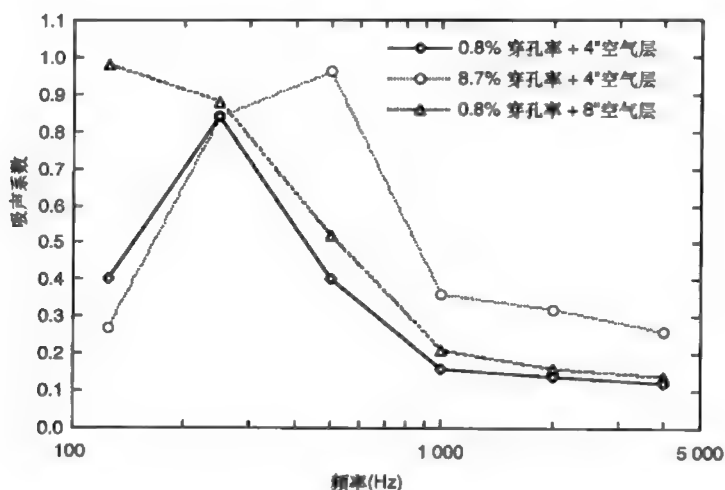


图 6.47 赫姆霍兹共振吸声结构典型吸声特性曲线

6.3.4 宽带吸声结构

将多孔吸声材料的吸声特性与共振式吸声结构的吸声特性结合起来，可以构成宽带吸声结构。典型结构如图 6.48 所示，其吸声特性如图 6.49 所示。所有使用了岩棉或玻璃纤维的吸声结构，都要注意防止恼人的纤维从吸声结构进入室内。

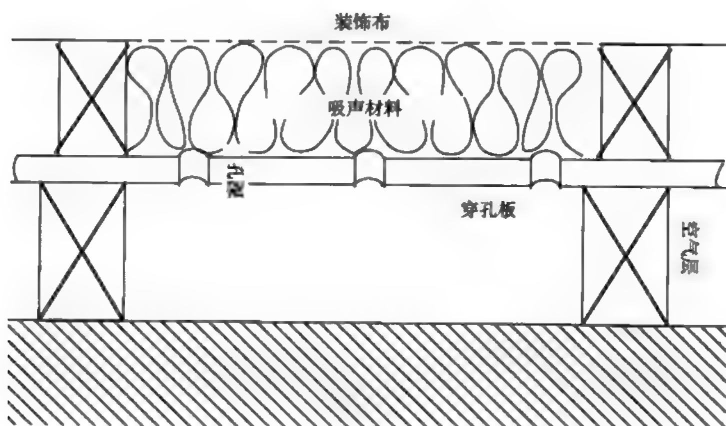
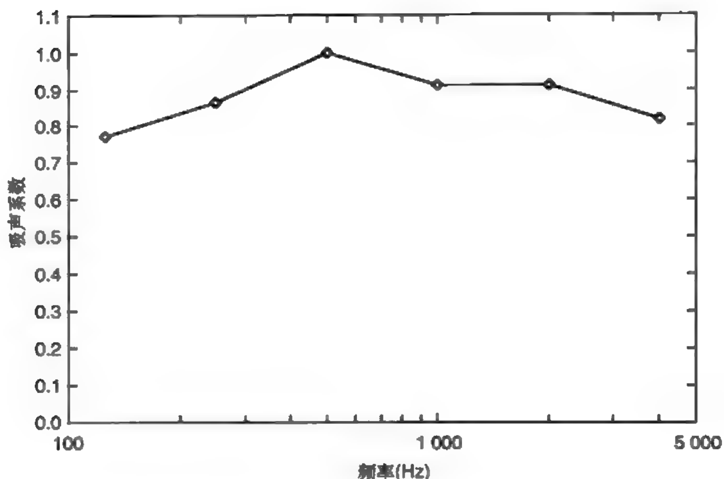


图 6.48 宽带吸声结构

图 6.49　宽带吸声结构典型吸声特性曲线



另一种获得宽带吸声特性的方法是采用很厚的多孔吸声材料，例如 1m 厚的多孔吸声材料，也可以获得宽阔平坦的吸声频率特性，但是造价却相当高。

6.3.5　小结

利用上述介绍的吸声材料和吸声结构，可以有效控制房间的吸声系数，使其具有所需的频率特性。在许多情况下，利用房间本身的材料就可以满足大部分吸声要求。例如，大多数巴洛克时期的音乐是在宅第的厅堂里演奏的，具有极好的声学平衡感。这是因为房间里使用了大量的面板装饰，这些木板可以有效地吸收低频，并与房间的另外一些装饰物如旗子、窗帘和壁毯等相结合，可以很好地满足吸声要求。

6.4　扩散结构

当声波撞击墙面时，一部分能量会被壁面吸收，然而同样基本的声学要求是声波被扩散反射。理想情况下，我们希望墙面在声学上是粗糙的，但不幸的是，大多数墙面，包括面积较大的吸声材料，对声波的作用更像是声学镜面。为了得到非镜面墙，可以将墙做成凹凸不平的形状。

有许多方法可用来达到此目的。凹凸的线度要达到至少 1/8 波长，最好是 1/4 波长，才能有效地使声波扩散反射。这就使得低频时要求物体的尺寸极大，如在 34Hz 时为 1.25m ~ 2.5m，而高频时所需的物体尺寸又很小，如在 3.4kHz 时为 1.25cm ~ 2.5cm。如果物体的尺寸太小，如小于 1/8 波长，就不能很好地扩散声波；如果物体的尺寸太大，如大于 1/2 波长，其本身就是声学上的镜面，因此也不能很好地扩散声波。可见，使声波扩散反射并不是轻易就能做到的。凹凸不平的表面对中频和高频有扩散作用。对于极高的频率，例如频率高于约 4kHz 时，表面粗糙的材料如砖块或切割粗糙的石头墙就能起到扩散作用。为了获得较好的扩散特性，一种“井”式结构的扩散体被研制出来并得到应用（Schroeder 1975、D'Antonio 和 Konnert 1984）。这种结构的“井”深是由严格定义的数值序列决定的。这些结构的设计原理相当复杂，如果读者想了解更详细的内容，可参看参考文献中所列的文献资料。下面将简要介绍它们的工作原理。

6.4.1 扩散体工作原理

假设某个坚硬表面上有许多高度为 d 的凸起物，并假设某声波从法向入射，那么，声波反射的状态与凸起物高度相对于波长的大小有关。下面分 3 种情况考虑：

（1）当 $d<<\lambda$ 时，表面可看成是平面，镜面反射声波；

（2）当 $d=\lambda/4$ 时，从表面最高点反射的声波比从表面反射的声波早 $\lambda/2$，这意味着在法向上声波相互抵消，没有声波在这个方向上传播。但是，由于系统不存在能量损失（能量守恒定律），因此声波必然向其他某个方向反射。如果声波不从法向入射，那么凸起物与平面的相对距离减小，镜面反射的声压幅度就会有所增大。这就是利用硬表面扩散声波的基本原理，即扩散面修正了反射波的波阵面相位或波阵面形状，使声波沿着非镜面反射方向反射；

（3）当 $d=\lambda/2$ 时，从表面最高点反射的声波比从表面反射的声波早 λ，因此入射波和反射波同相。这时，凸起物的影响消失，硬表面可看成是平面，对声波产生镜像反射。

由此可见，有规则的凸起物能够扩散声波，但是只对 $\lambda/4$ 奇数倍频率有效，而且与声波的入射方向有关。但是，我们所需的能够改变反射波波阵面相位的凸起物结构应满足下面两个物理条件：

（1）声波以“最佳”方式扩散；

（2）在足够宽的频率范围内产生最佳扩散。

可以通过一些不同高度的序列来满足上述物理要求，但是它们必须具备以下两个共同特性：

（1）序列的傅里叶变换为均匀恒定值，只是直流分量可能相同或者较低。这一点满足了第一个要求，事实证明这种性质的反射面能够向各个方向有效反射声能。较小的直流分量有利于进一步减小镜像反射的能量。

（2）这些序列应具备的第二个性质是，当入射波波长发生变化时，傅里叶变换不受影响。这会导致序列的阶次发生变化，但是可以看到它们仍然具有原来傅里叶变换的特性。

这些序列之所以要满足上述两个性质，是因为序列的作用是在整个波长周期内干扰波阵面。由于这种序列干扰了波阵面的相位，因此这类扩散体称为相位反射栅（phase reflection gratings）扩散体。

下面分析两个用于构成扩散体的具体序列，来进一步解释上述问题。

（1）二次剩余序列：“井”深为 n^2 除以 p 的余数，其中 p 为质数。如果取 $p=5$，则得到“井”深序列为 0，1，4，4，1，0，1…（取 $n=0$，1，2…$n-1$）。

可见，该序列的周期为 5 个数。

（2）原始根序列：“井”深为 a^n 除以 p 的余数，其中 p 为质数，a 为一个恰当的常数，称为原始根。如果取 $a=2$，$p=5$，则序列为 1，2，4，3，1，2……（取 $n=0$，1，2…$n-2$）。

此序列的周期为 5−1＝4。

在上述例子中，最低设计频率由“井”深 5 等于 $\lambda/2$ 决定。对于更高的频率，这些序列能够保持相同的性质并有效扩散声波。但是，当频率足够高使得 $\lambda/2$ 等于“井”深的最小差值 1 时，扩散面将重新等效为平面。

这种扩散体的典型结构如图 6.50 所示，其扩散性能如图 6.51 所示。

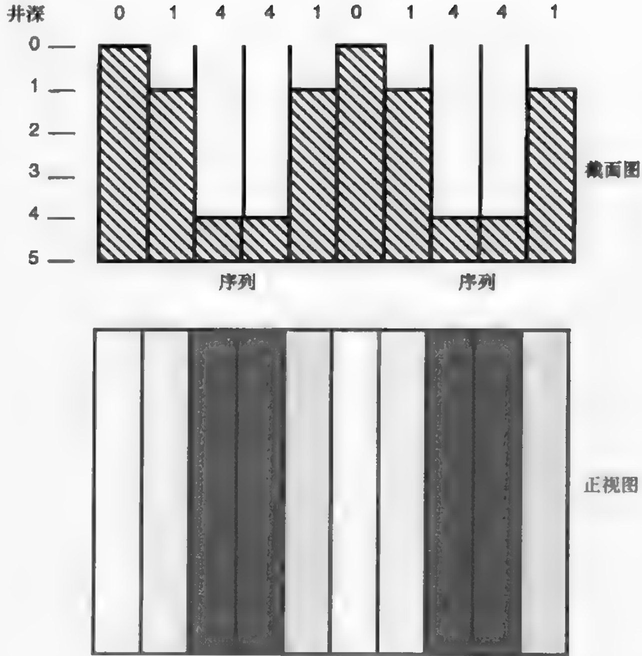


图 6.50 典型二次余数扩散体结构

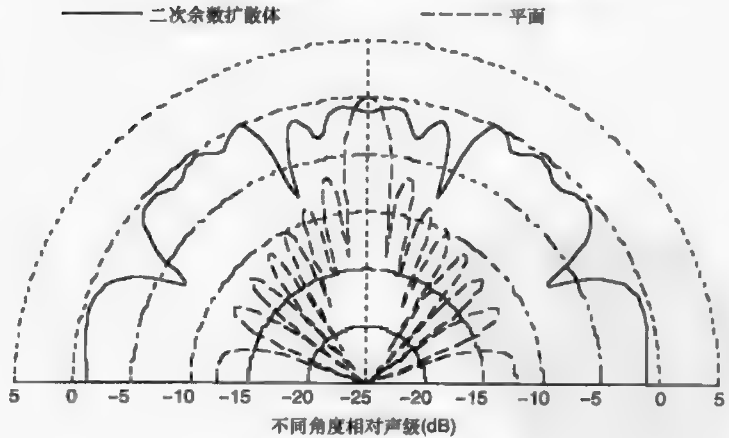


图 6.51 二次余数扩散体与平面反射的典型性能比较（注意为方便起见将镜像反射设为 0dB。实际平面反射在该方向的输出更大。）

6.4.2 讨论

正如我们所看到的，这些序列通过在一个周期范围内改变反射波波阵面的相位来获得其扩散性能。为了达到此目的，其最大值应该等于最

低设计频率的二分之一波长（λ/2），这意味着为了满足扩散要求，合理的“井”深设计是必要的。例如，如果最低设计频率为 500Hz，则井深需要达到 34cm(13.5insh)；如果设计频率降低到 250Hz，则井深需要增大一倍。然而我们已经看到，一种简单的 λ/4 凸起结构就能够对特定频率起到扩散作用，而这个深度只是上述最大值的一半，因此它可以代表扩散体的有效工作上限频率。

如果将最大井深设为接近于最低频率波长的四分之一（λ/4），也可能获得良好的扩散特性，这对于较小空间是一个好的解决办法。然而，即使选择 λ/4 的井深，仍然使扩散体占用的体积太大而不实用。我们所希望的扩散体最好没有任何深度！

6.4.3 幅度反射栅扩散体

并不仅仅是肉眼可见的墙面凸起物能使声波扩散反射，实际上表面反射特性的变化也会引起扩散反射。例如墙面的吸收特性从吸声变化到反射就能引起声波的扩散反射。因此，为了改善扩散性能，最好将吸声材料随机分散地安装在墙面各处，而不是简单地将其安装在某个特定区域。除了有利于扩散外，这种安装方式还能避免形成两个吸声较弱的平面之间的颤动回声。如何分配吸声材料以获得最佳扩散效果也存在一些数学依据，详细内容可参看 Angus(1995) 的文献。实际上我们需要的是一种幅度加权，即形成一种不同吸声特性的排列，使之具有均匀平坦的傅里叶变换。

最容易想到的序列是二进制数字序列，即它们仅由 0 和 1 构成，其中 1 表示硬反射面，0 表示某种吸声材料构成的吸声面。由于没有一种吸声材料能够达到 100% 的吸声效果，因此只要将序列中的 0 用 $1-\alpha$ 代替即可。吸声系数不能达到 100% 的后果是使镜像反射部分的能量增加了。在许多可行的二进制数字序列中，m 序列（译者注：最大长度序列，即 MLS 序列）是最好的选择，因为它们具有所需的傅里叶变换性质。还有许多其他的两个数组成的序列，它们也有平坦的傅里叶变换，但是只有 m 序列是被充分论证的。

幅度反射栅（amplitude reflection gratings）是指经过声学处理的表面。

它实际上是由一定宽度的吸声性质的板条和反射性质的板条按一定形式排列构成，吸声板条的宽度小于最高频率波长的一半（λ/2），代表数字 0，反射板条代表数字 1，如图 6.52 所示。值得注意的是，因为这种扩散体并不是靠深度对声波产生扩散的，因此不存在低频下限频率，而高频上限频率是由板条的宽度决定的。

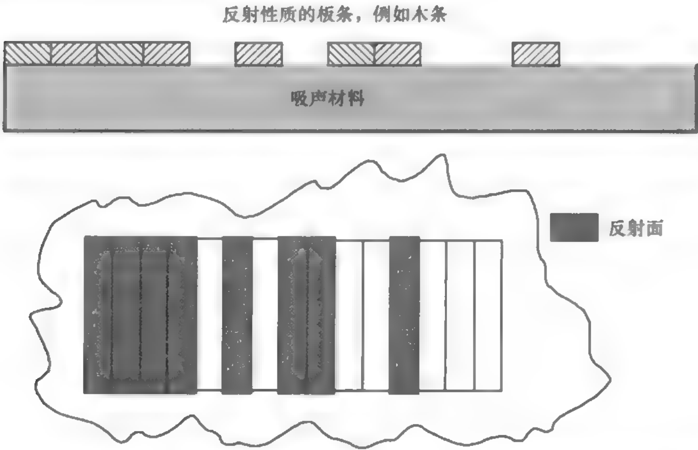


图 6.52 周期为 15 的一维幅度反射栅示意图

图 6.53 所示为二维幅度反射栅的构造示意图。虽然幅度反射栅没有

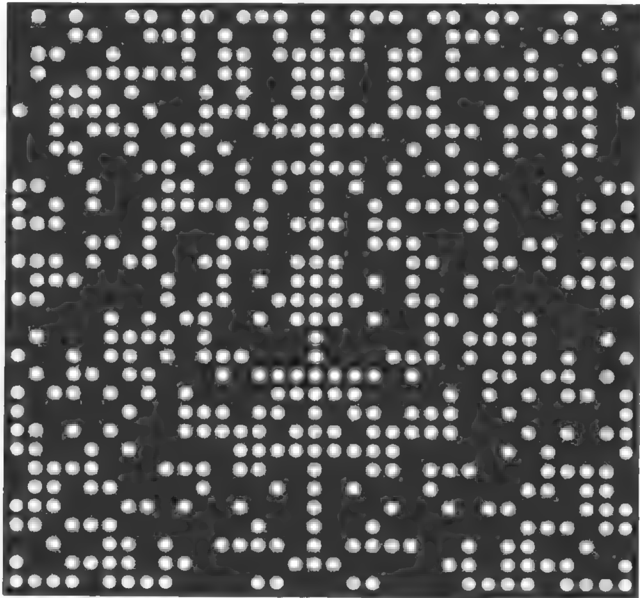


图 6.53 周期为 1 023 的二维幅度反射栅扩散体结构示意图（白点表示吸声面上的洞，代表反射面）

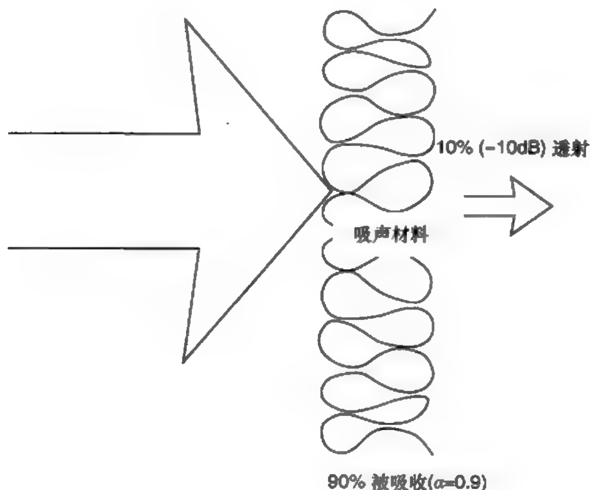
相位反射栅那样好的扩散性能，但它对声波有一定的扩散反射作用。由于它不像相位反射栅那样占用许多空间体积，因此是一种有效的低频扩散体。幅度反射栅也可以制成曲面结构，但没有数学依据。这方面的详细内容可参看 Cox（1996）的文献。还有其他结构的扩散体，读者可参看参考文献中所列的文献资料。

6.5 隔声

如果没有讨论怎样将外面的噪声进行隔离或使房间内的声音不会泄露到外面，使房间内外的人都满意，那么有关房间音质控制的讨论是不全面的。

首先要认识到，并不是有较好吸声性能的某种材料，它也能有效隔离声波。事实上，大多数吸声材料的隔声性能都很差，这是因为在研究隔声性能时，我们只对有多少声能穿过材料这个问题感兴趣，而不是对有多少声能被吸收感兴趣，如图 6.54 所示。20dB 左右的隔声量表示很差的隔声效果，虽然它相当于只有 1/100 的声能穿透过去。吸声系数为 0.9 的良好吸声体意味着允许 1/10 的能量穿过，相当于隔声量仅为 10dB！ 良好的隔声要求隔声量至少为 40dB，显然吸声材料不能作为隔声使用！

图 6.54 声透射与声吸收的区别



6.5.1 隔声方法

通常只有两种方式可以达到隔声目的，即通过弹性和质量。图 6.55 所示为某一隔断的隔声频率特性。由图可知，由于声波必须作用于弹性隔断上，因此低频时弹性对隔声起主要作用，这一区域称为弹性控制隔声区。随着频率升高，隔断辐射一定声压级所需的振动位移减小，因此隔声效果越来越差，直到频率达到隔断的共振频率时，隔声量达到最小值。在共振时，较小的声波就能使隔断产生较强的振动，因此隔声效果最差。当频率高于隔断的共振频率时，系统进入质量控制区。在此区域，声波要使沉重的质量加速运动，因此质量对隔声起根本性的作用。由于高频时移动质量所需的作用力较大，因此隔声量随频率升高而增大。当频率更高时，入射声波与隔断的厚度、隔断内部波动方式发生相互作用，产生了吻合共振，使隔断的隔声效果减弱。阻尼可用来抑制吻合共振，从而减小吻合共振带来的这种影响。大部分隔断工作在隔声曲线的质量控制区，并以吻合共振频率作为工作频率上限。图 6.56 所示为几种单层隔断的隔声频率特性曲线。特别值得注意的是，石膏板墙的吻合共振效应较显著。对于单层隔断而言，质量加倍，则隔声量增大约 3dB，同时吻合共振频率也会有一定程度的降低。吻合效应限制了单层墙的高频隔声性能。此外，用单层隔断实现强隔声的造价较高，尺寸较大也使其不具实用性。

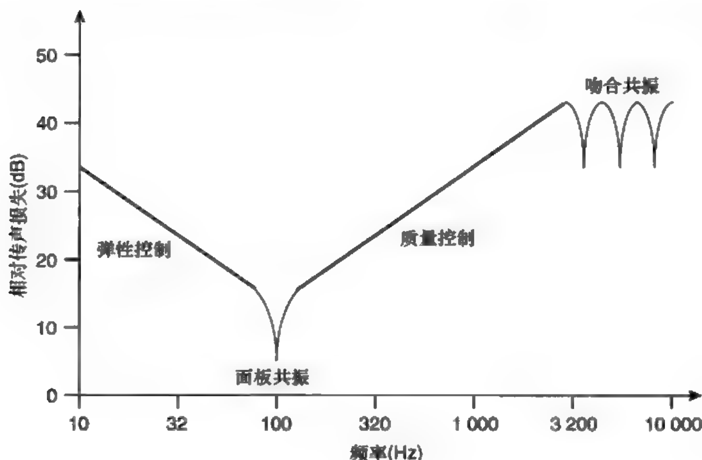


图 6.55　某隔断的隔声频率特性

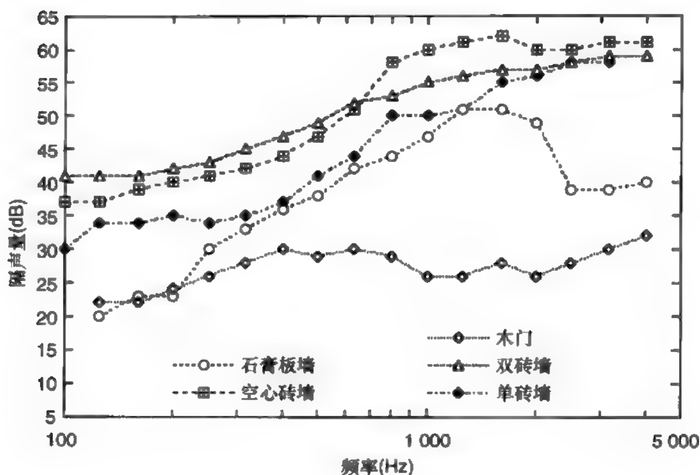
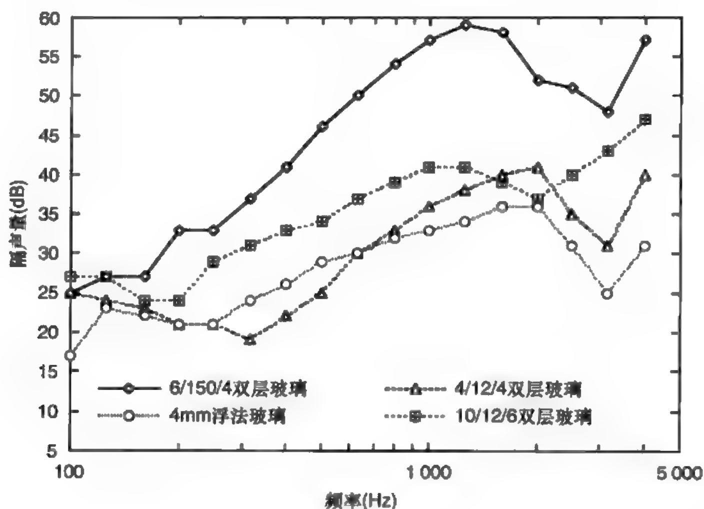


图 6.56 几种常用隔断隔声频率特性

6.5.2 独立隔断

解决强隔声的办法是同时使用两个或更多的相互独立的隔断。如果两个隔断真正是相互独立的，那么其总的声衰减或有效隔声量是各个隔断声衰减的乘积，或衰减分贝值为各个独立隔断的分贝值之和。在实际应用中，由于不同隔断层之间不是完全独立的，虽然隔声量大大提高了，但并不像所预测的那样为不同隔断隔声量的简单相加。吻合共振会减弱隔断的隔声效果，因此，要使它们各自有不同的吻合共振频率，这一点是很重要的。要达到这个目的很简单，只要使材料的厚度不同或选用不同的材料即可。图 6.57 所示为来自 Inman(1994) 的几种单层和双层玻璃的隔声量测量结果，其中双层玻璃的空气间隙和厚度各不相同。由于吻合共振，厚度同为 4mm 的双层玻璃的隔声效果实际上比 4mm 厚的单层玻璃差！另外两条曲线则说明，当两层玻璃厚度不同时，情况会大大改善。当增大两层玻璃之间的空气层从而减小两层玻璃之间的耦合程度时，隔声量会进一步提高。通常在两个隔断之间会放置一些吸声材料，用于减小吻合共振效应。但需要注意的是，吸声材料和两个隔断之间不能有接触，否则会由于结构性传导使隔声效果减弱。

图 6.57 几种单层和双层玻璃隔声频率特性



6.5.3 结构性传导路径

声波的结构性传导是隔声效果减弱的主要原因。当声波能够绕过隔断传播时，这些传播路径都称为结构性传导路径，如图 6.58 所示。常见的结构性传导路径有建筑物框架结构、暖气管道、通风管道和空隙传导。建筑物框架的结构性传导可以通过建造“屋中屋”或称为浮筑结构来减小，如图 6.59 所示。它是通过将房间建立在弹簧之上来实现房间与框架结构之间的隔离。在实际应用中，要使浮筑房间与建筑结构框架完全隔离是极为困难的（管道装置和线管往往是二者之间产生连接的原因）。在许多情况下，通风系统和缝隙也是产生结构性传导路径的主要原因。事实上，在非专业领域，声音的隔离度几乎完全由缝隙和通风系统决定。双层玻璃窗的推销商经常通过消除这些声音传导路径后，夸耀其产品极好的隔声性能，甚至这些双层窗可能是由两片厚度同为 4mm 的玻璃构成。

因此，为了获得良好的隔声效果，人们需要一个具有优良性能的隔断、无缝隙、无通风系统的房间结构，显然这样的房间是不切实际的，因此很难做到。

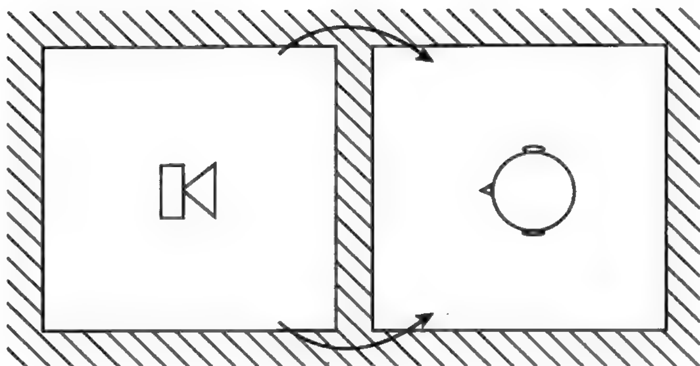


图 6.58　结构性传导路径示意图

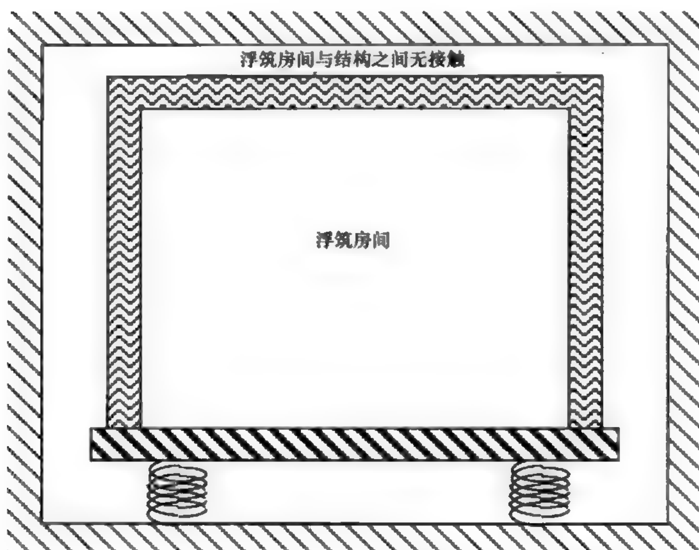


图 6.59　浮筑房间结构示意图

6.6　声能时间特性的考虑

在听音室的声学设计中，除了要考虑混响时间外，还要考虑房间声能建立的时间过程。现在已经可以利用声学测量系统对房间声能的时间特性直接进行测量，这样设计者可以看到房间的声学性能，而不是仅仅依靠“金耳朵”来判断声音的好坏。理想的能量时间特性如图 6.60 所示，它主要有以下 3 个特征：

（1）直达声和第一个反射声的时间间隔。大多数房间存在这样的时间间隔，它主要给听音者房间大小的暗示。这个时间间隔不能太长，一般要求小于 30ms，否则容易产生回声。同时，一定的延时是需要的，它使直达声占据一定的时间空间，从而改善声音的清晰度。此外，较小的时间间隔往往能增加声音的"亲切感"。

（2）存在较强并且扩散的早期反射声。这些反射声主要从侧面方向到达听音者，称为侧向早期反射声。它能够增加声音的空间感，往往在鞋盒式厅堂比在扇形厅堂更容易在整个听众区获得侧向早期反射声。第一个早期反射声应该在直达声后 20ms 之内到达听音者。理想的早期反射声的频率特性应为平直的，加之需要较强声级的早期反射声，意味着厅堂的侧墙应为扩散面，并且吸声系数应尽量小。

（3）一个平稳衰减的扩散混响声场，没有明显的缺陷，没有模式共振，具有与所表现的音乐相适宜的混响时间。在实际应用中很难做到这一点，在大多数情况下需要折衷考虑，并在某些方面做出妥协。对于声学乐器演奏的音乐而言，在混响中有一定的低频提升可以增加声音的"温暖感"。但是，如果是在演播室中，这种低频提升并不是那样有必要。

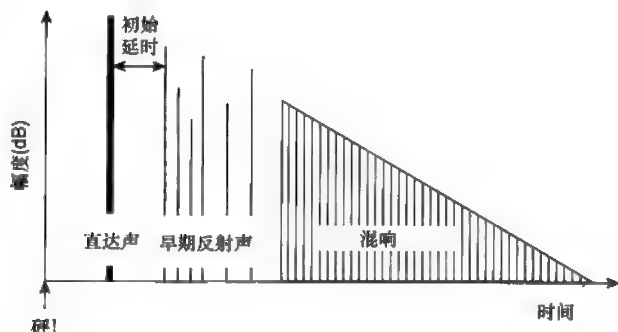


图 6.60　理想的声能衰减特性

6.6.1　"无反射区"听音室

上述声能衰减条件适用于音乐厅的声学设计，在某种程度上，也适合于音乐录音棚中演奏者所在区域的声学设计。然而，对于家庭听音室

或录音控制室而言，理想的情况是听音者能够透过重放系统听到录音棚里原来的声音。遗憾的是，一般重放录音作品的房间都比录音棚小得多，因此产生了图 6.61 所示的声音效果。这里听音者所听到的第一个反射声是来自听音室，而不是来自录音所在的声学环境。根据优先效应，这个反射声将占主导作用，听音者感到重放声来自听音室大小的空间，这显然不是所希望的。因此，为了使扬声器发出的声音好像来自一个较大的空间，需要抑制这些来自听音室侧墙的反射声，如图 6.62 所示。达到这一目的的方法之一是采用吸声措施，如图 6.63 所示，也可以通过使侧墙倾斜一定角度来达到此目的。这种设计方法称为 "无反射区(reflection-free zone)" 技术，它的基本设计思路是消除听音室某个区域的早期反射声，从而获得较大的初始延时。这种听音效果只能在有限的区域内获得，除非把整个房间设计成消声室，这当然是不行的。这种设计思路是将来自所有墙壁的第一次反射声吸收或者导向其他区域，但允许高次反射声存在，从而使初始延时最大化。如果时间间隔大于原始录音棚的初始延时，听音者就能够感受到录音棚大小的空间，而不是听音室的小空间。在进行这样处理的同时，还要注意满足均匀扩散混响声场的要求，为此通常将后墙设计成扩散性能良好的扩散反射面。听音位置的初始延时应尽可能大，但注意不要大于声音从扬声器到后墙再反射到听音者的时间间隔。理想情况下，初始时间间隔大约为 20ms，决不能比这个值大很多，否则可能产生回声。在实际应用中，这个条件一般都能得到满足，初始时间间隔大约在 8ms ~ 20ms。

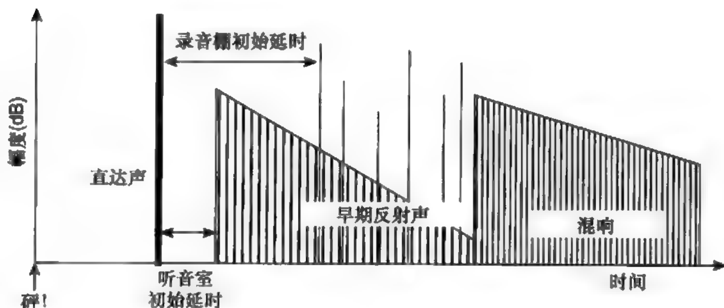


图 6.61　听音室较短初始延时对听音的影响

图 6.62　抑制早期反射声使初始延时最大化

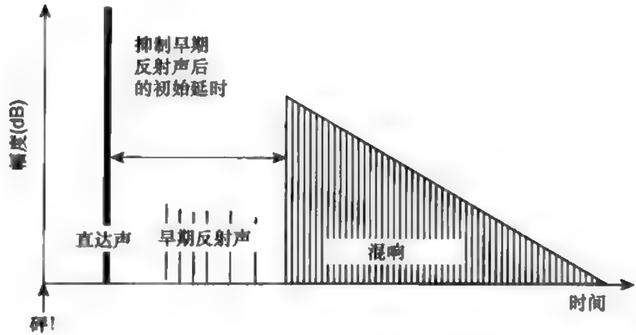
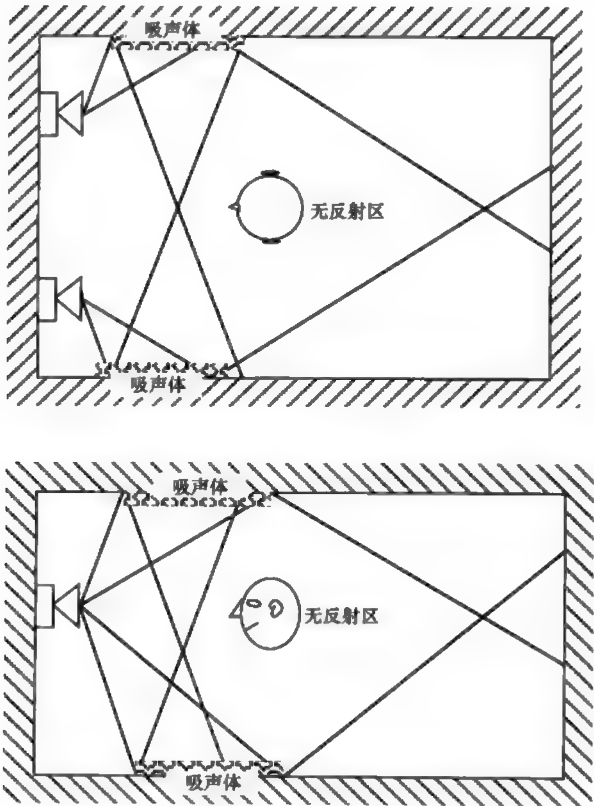


图 6.63　通过吸声获得无反射区



6.6.2　无反射区对吸声的要求

为了获得无反射区，需要抑制早期反射声，但是在多大程度进行抑制呢？图 6.64 所示为早期反射声干扰立体声声像定位的平均阈值。由此可知，反射声的声级必须比直达声小大约 15dB，才能对主观听音不产生影响。考虑到声波传播的平方反比定律，允许反射声的阈值有一定的提升，也就是说，反射声的相对声级只要小于约 -10dB 即可，这就意味着产生反射声的相应墙壁的吸声系数应达到约 0.9。在家庭听音室，可以通过挂毯和窗帘近似这样的吸收要求。书架可以作为有效的扩散体。如果要说服某些房主承认用地毯或窗帘装饰天花板是很优雅的，这恐怕难以做到！在演播室里，可以采用较为极端的声学处理方式，但同时还要注意使整体的声学效果保持良好，并且听感舒适，决不能做成消声室。还要注意到，由于声频的波长范围较大，这种处理方法只对中、高频有效，因为只有在中、高频段，这种局部处理的尺寸才可以与波长相比拟或比波长大得多，才能满足镜像反射原理。

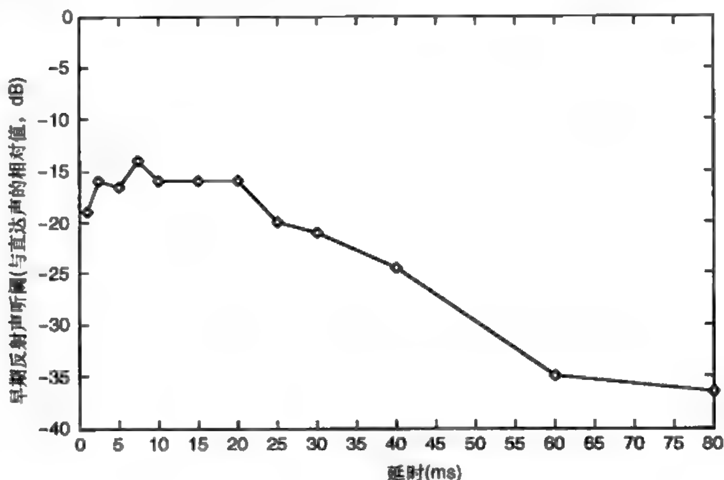


图 6.64　无反射区所需的反射声抑制量（Toole 1990）

6.6.3　“无环境”听音室

另一种成功应用于许多听音室的早期反射声控制方法称为“无环境

(non-environment)”听音室。它既控制早期反射声，也控制混响声。虽然在声学意义上这类听音室非常沉寂，但是它们不是消声室。这类听音室的声学处理如图 6.65 所示。它的扬声器嵌入式安装在反射面里，地面通常是反射面，而后墙是强吸声处理，两侧墙也是强吸声处理或部分的强吸声处理。这样处理的结果使来自扬声器的声音被吸收而不是被反射，听音者除了听到一些来自地面的反射声外，只能听到扬声器的直达声。而对于远离扬声器的其他虚声源即一次发射声，房间里的两个反射面又能提供一定的混响声。这就是说，对于房间里的听音者而言，虽然声学环境是沉寂的，但又不像消声室那样使人感到压抑。这类听音室的倡导者认为，当只有来自扬声器的直达声而不存在任何其他声音时，听音者能够清楚地听到重放声中较低电平的声音细节，产生极好的立体声声像定位。这主要是因为去掉了对声像定位有影响的所有反射声，而来自地面

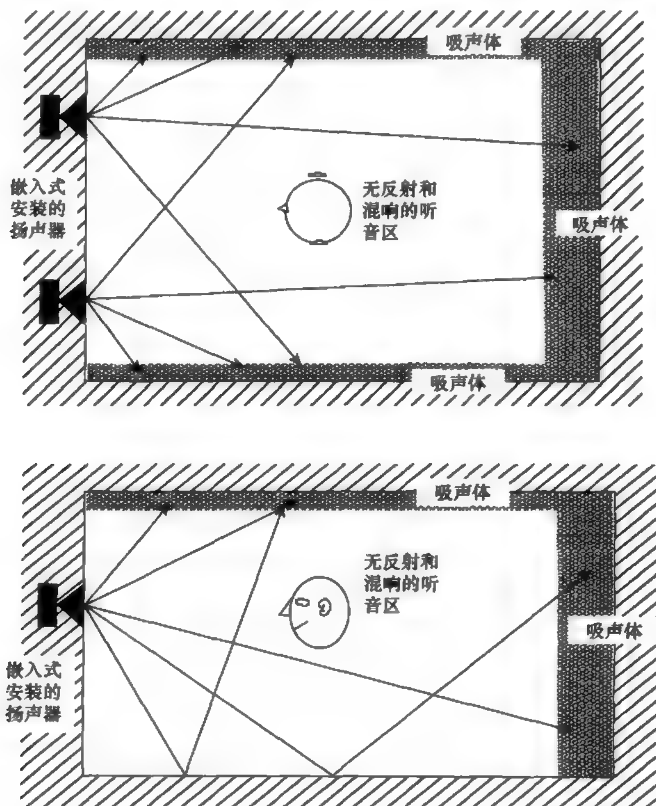


图 6.65 “无环境”听音室

的反射声对声像定位几乎没有影响。理论上要求这类听音室使用宽带吸声材料，以及能够不借助房间的反射产生所需重放声压级的专业扬声器系统。有关这类听音室设计的更多资料可参看 Newell（2003）的文献。

6.7　边界对扬声器输出频率特性的影响

边界如墙壁等对扬声器低频输出的提升效应是众所周知的（Allison 1980）。许多方法可以用来改善扬声器的输出频率特性，减小边界的影响，例如，对扬声器进行合理摆位，或对扬声器进行针对性设计等。

当扬声器放置在某个反射面附近时，会产生一个镜像声源，如第 1 章中的图 1.24 所示。因此，当扬声器放置在 3 个边界面附近时，会产生 3 个镜像声源，如图 6.66 所示。这些镜像声源代表相应界面的反射声，对扬声器的辐射阻抗将产生影响。由于电动式扬声器在工作频率范围是质量控制型，因此可看成一个高阻抗源，其辐射声功率随着辐射阻抗的变化而变化。

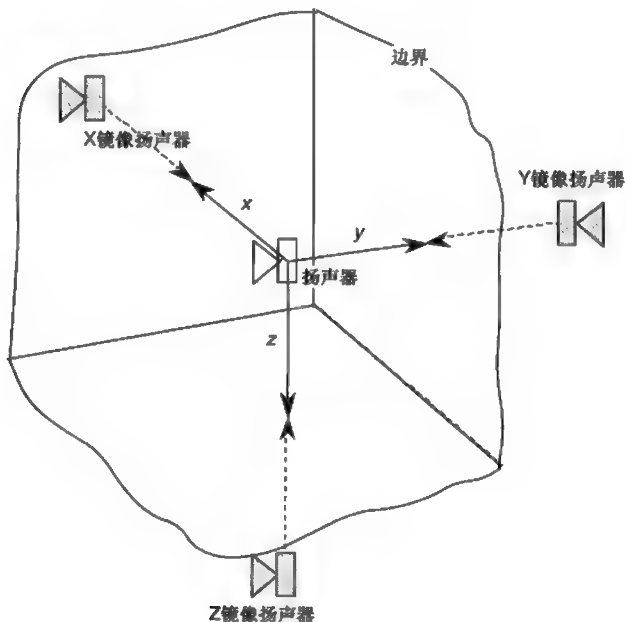


图 6.66　离扬声器最近的 3 个界面形成的镜像声源

Waterhouse（1955）对 Rayleigh（1964）的研究工作进行了拓展，通过实验证实了声源的声输出受边界如房间墙面存的影响。边界对声源输出的影响可用图 6.67 和图 6.68 所示的曲线表示，分别代表两种不同的边界条件。其中一种情况是所有方向到边界的距离相等（图 6.67）；另一种情况是到各个边界的距离各不相等，使边界的影响最小化（图 6.68）。到某个特定界面的距离并不重要，重要的是到各个界面距离的比例关系。还要注意到图中 *x* 轴表示距离 "*x*" 所包含的波长数目，当 *x* 值不变时，这个变量相当于频率，只不过引入距离作为参考值。因此，如果 "*x*" 等于 1m，则 *x* 轴上的 1 代表波长为 1m（频率为 344Hz）；如果 "*x*" 等于 0.5m，则 *x* 轴上的 1 代表波长为 0.5m（频率为 688Hz），依此类推。

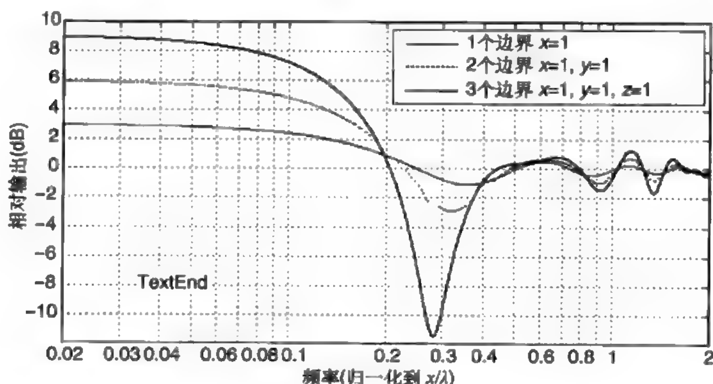


图 6.67　非最佳摆位时界面对扬声器声输出的影响

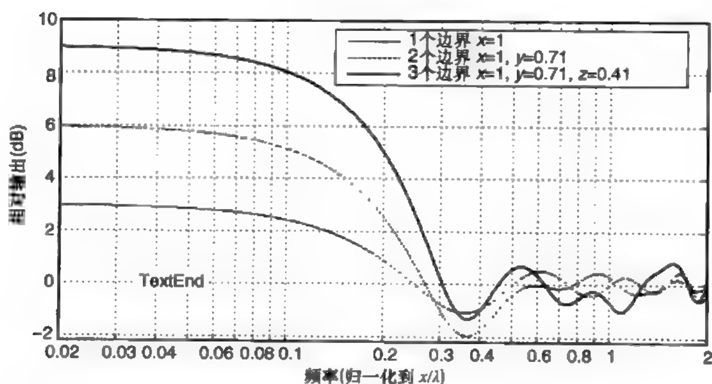


图 6.68　最佳摆位时界面对扬声器声输出的影响

从这两组曲线可以得到以下两个主要结论：

（1）如果到界面的距离都相等，那么低频响应要比中频响应高约

9dB，在整个频率范围频响起伏较大，特别是当距离为 1/4 波长时，频响出现一个明显的谷值。

（2）通过调整扬声器到各个界面的距离，可以消除某些频响的起伏变化，但 9dB 的低频提升仍然存在。

上述结论是大家所熟知的，而且 9dB 的低频提升在 3 个界面同时存在的情况下是固有的。一般通过对放置在 3 个界面附近的扬声器的频率特性进行适当设计，可以对界面产生的这种效应进行补偿。

6.8　箱体衍射效应的减小

箱体衍射效应的存在也会对扬声器的输出产生影响。Olson(1969) 于 1950 年用实验方法演示了箱体的形状和大小对扬声器频率响应的影响。在波长与箱体尺寸相当或小于箱体尺寸的任何频率，都会发生衍射效应。这就是说，即使是很小的箱体，对于安装在顶部的高音单元，也会发生衍射。对于主箱体，只要扬声器单元工作频率对应的波长与箱体尺寸相当，衍射效应就会发生。

在设计扬声器时，了解扬声器辐射的声波要经历一个从有边界到无边界的传播过程是很重要的。这就是扬声器发出的声波沿着箱体的面板传播时所发生的情况，这时，声波的状态从被面板限制转换到不受任何限制，如图 6.69 所示。其产生的反射声与主输出发生干涉，使扬声器的频率响应变得不均匀。

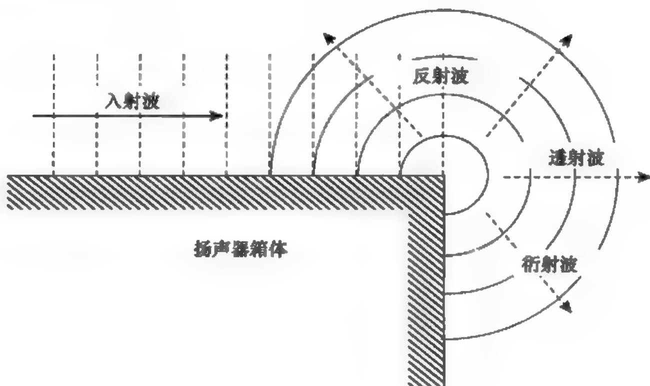


图 6.69　声波在扬声器箱体边缘从有界状态到无界状态引起的反射

Olson 的部分实验结果如图 6.70 所示。他的研究表明，理想的箱体形状应为流线型的，尺寸应比工作频率对应的波长小。他同时指出，频响曲线最为光滑的箱体形状为球形箱体，如图 6.70(b) 所示。他还指出，一些非球形箱体，例如锥台和直角平行六面体的组合体，如图 6.70(c) 所示，也能提供良好的频率响应。

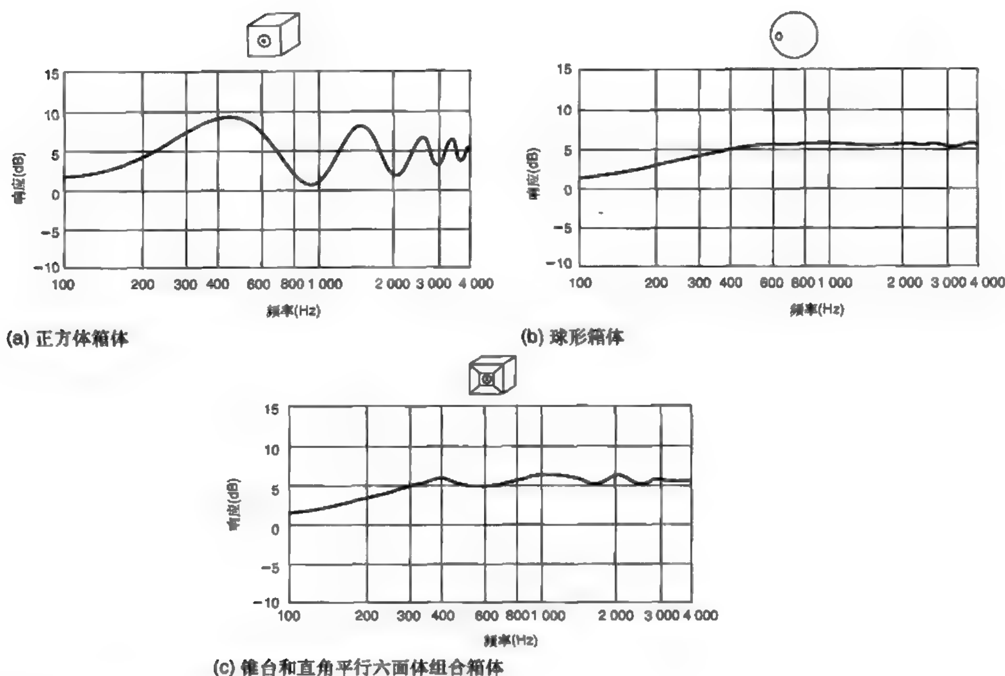


图 6.70 不同箱体产生的衍射效应（Olson 1969）

虽然图 6.70(b) 所示的球形箱体是最为理想的，但是其他曲线形的箱体却具有某些实际应用的优势。例如，图 6.71(b) 所示的截去头部的椭圆形箱体不仅具有良好的衍射性能，而且能够在面板面积一定的情况下提供灵活的箱体体积。这一点是非常重要的，主要基于以下几个理由：

（1）为了基本上保持球体形状，球形箱体的直径一般要比扬声器单元直径大很多。在图 6.71(a) 中，球体的直径大约是扬声器单元直径

的 2 倍。如果要使球体的直径等于扬声器单元的直径，那么箱体的形状就是半球形，其性能就会变差。

（2）箱体的体积对扬声器的性能有一定的影响，与扬声器共同决定音箱的工作频率下限。球形箱体的直径不可能小于扬声器单元的直径，而图 6.71（b）所示的椭圆形箱体，可以不依赖于扬声器单元的直径在一定程度上自由选择箱体体积。

（3）球形箱体容易产生重叠共振，影响扬声器的频率响应和时域响应，而图 6.71（b）所示的椭圆形箱体不存在重叠共振，因此其频率响应更为光滑。

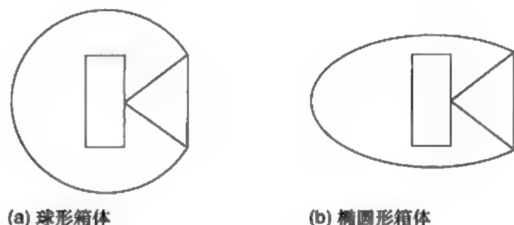


图 6.71　球形箱体和椭圆形箱体的比较

Olson 的研究结果说明，椭圆形箱体也可以达到和球形箱体同样好的频率响应。

6.8.1　局部为平面的箱体衍射效应

由于低频的波长大于箱体的尺寸，因此低频的衍射现象并不严重。但是，如果主箱体包含中频扬声器单元，那么仍然需要考虑衍射问题。通常箱体上还同时安装有低频单元，而且箱体必须放置在平面上，因此不可能使用球形箱体，曲面只能用于沿着垂直轴线的方向上。Olson（1969）也对这类箱体形状进行了测试，所得到的部分结果如图 6.72 所示。从图 6.72（a）可以看出，矩形箱体会产生严重的衍射效应，即使扬声器安装在面板非中心位置，避免了产生重叠共振，但这种衍射效应仍然十分显著。图 6.72（b）所示的圆柱形箱体似乎也不能起到很大的改善作用，这可能是因为两者都存在两个箱体边缘以及由于扬声器单元置于这两个边缘中心位置所引起的重叠共振现象。如果扬声器单元安装在面板的非

中心位置，会改善其频率响应。图 6.72(c) 所示为梯台和矩形组合成的箱体，并且扬声器单元安装在非中心位置，其频率响应大大得到改善。如果面板是光滑的曲面而不是平面，频率响应会进一步得到改善，这时，对声辐射的干扰主要来自箱体背板界面的突变。这些研究结果表明，为了改善扬声器的频率响应，扬声器单元附近的面板应尽量采用光滑的曲面，并且扬声器单元应安装在面板的非中心位置。在实际应用中，一般扬声器单元在水平方向要安装在中心位置，而在垂直方向可以不安装在中心位置。在这种情况下，最合适的箱体形状是球形，但是，由于前面所述的原因，椭圆形箱体常常更为合适。

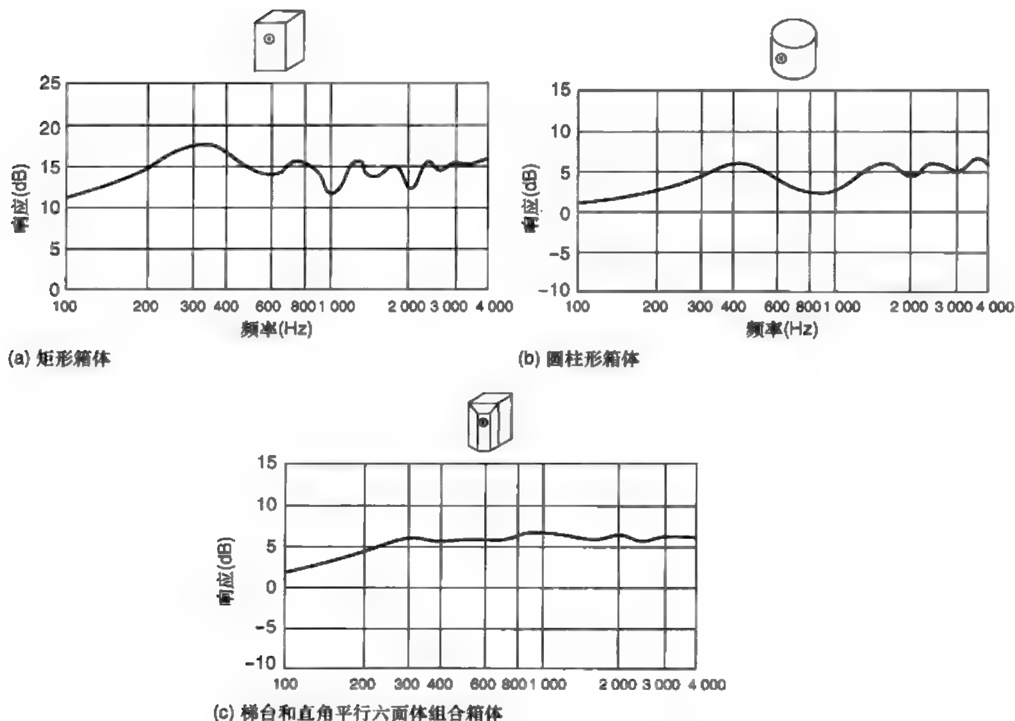


图 6.72 至少包含两个平面的箱体衍射效应（Olson 1969）

本章探讨了重放声音的声学环境如何影响声音的重放和声音的听感，同时分析了各种不同的重放环境的声学特性，探讨了获得良好听音环境的几种声学处理技术。

参考文献

Allison, R.F. (1980). The influence of room boundaries on loudspeaker power output, in Loudspeakers an Anthology Vol. 1, ed. Cooke R.E. New York: The Audio Engineering Society, pp. 353–359.

Angus, J.A.S. (1995). Sound diffusers using reactive absorption gratings. Audio Engineering Society 98th Convention, 25–28 February, Paris, preprint #3953.

Angus, J.A.S. and McManmon, C.I. (1998). Orthogonal sequence modulated phase reflection gratings for wide-band diffusion. Journal of the Audio Engineering Society, 46, (12), December.

Cox, T.J. (1996). Designing curved diffusers for performance spaces. Journal of the Audio Engineering Society, 44, (5), 354.

D' Antonio, P. and Cox, T. (1998). Two decades of diffuser design and development, Part 1: Applications and design. Journal of the Audio Engineering Society, 46, (11), November.

D' Antonio, P. and Konnert, J.H. (1984). The reflection phase grating diffusor: design theory and application. Journal of the Audio Engineering Society, 32, (4), April, 228–238.

Holman, T. (1999). 5.1 Surround Sound. Boston: Focal Press.

Inman, C. (1994). A practical guide to the selection of glazing for acoustic performance in buildings. Acoustic Bulletin, 19, (5), September–October, 19–24.

Lord Rayleigh (1964). Work done by detached sources, in Scientific Papers. Dover, New York, vols. 5–6, pp. 135–141.

Newell, P. (1995). Studio Monitoring Design. Oxford: Focal Press.

Newell, P. (2000). Project Studios: A more professional approach. Oxford: Focal Press.

Newell, P. (2003). Recording Studio Design. Oxford: Focal Press.

Olson, H.F. (1969). Direct radiator loudspeaker enclosures. Journal of the

Audio Engineering Society, January, Vol. 17, Number 1, 22–29. Note, originally presented at the Second Annual Convention of the Engineering Society, 27 October, 1950.

Schroeder, M.R. (1975). Diffuse sound reflection by maximum length sequences. Journal of the Acoustical Society of America, 57, (1), January, 149–150.

Toole, F.E. (1990). Loudspeakers and rooms for stereophonic sound reproduction, in The Proceedings of the Audio Engineering Society 8th International Conference, The Sound of Audio, 3–6 May, Washington DC, 71–91.

Walker, R. (1993). A new approach to the design of control room acoustics for stereophony. Audio Engineering Society Convention, preprint #3543, 94.

Walker, R. (1996). Optimum dimension ratios for small rooms. Audio Engineering Society Convention, preprint #4191, 100.

Walker, R. (1998). A controlled-reflection listening room for multichannel sound. Audio Engineering Society Convention, preprint #4645, 104.

Waterhouse, R.V. (1955). Output of a sound source in a reverberation chamber and other reflecting environments. Journal of the Acoustical Society of America, 27, no. 2, March.

第 7 章　声音的电子加工和处理

在前面几个章节我们探讨了与音乐有关的声学和心理声学基本概念和理论，并且在某些特例中，强调了声学在演播室、剧院和乐器设计中的应用。然而，声音的电子加工和处理是现代音乐制作和生产的重要环节之一。因此，在最后一章，我们将从声学和心理声学角度介绍声音的电子加工和处理技术。

7.1　滤波处理

最简单的声音信号处理是滤波处理，即去掉不需要的频率成分。例如，我们经常需要把传声器拾取的低频噪声如通风系统产生的噪声和交通噪声滤除，这项工作可以由高通滤波器来完成。出于上述原因，调音台通常配备某种形式的高通滤波器。为了减小噪声和失真，或者为了滤除高频成分以防止在数字信号处理中产生混叠失真，也经常需要滤除高频成分，这可以通过低通滤波器来完成。第 3 种滤波器称为陷波器，常用于滤除声音信号中的某一频率成分。图 7.1 所示为这些不同种类滤波器对典型的音乐信号频谱产生的影响。

无论上述哪一种滤波器，理想的状态是信号经过处理后对主观听音

产生的影响最小化。虽然理想的情况是处理过的声音信号的音色没有改变，但实际上或多或少会对音色产生一些影响。这些影响是什么？如何利用声学和心理声学知识使这些影响最小化？

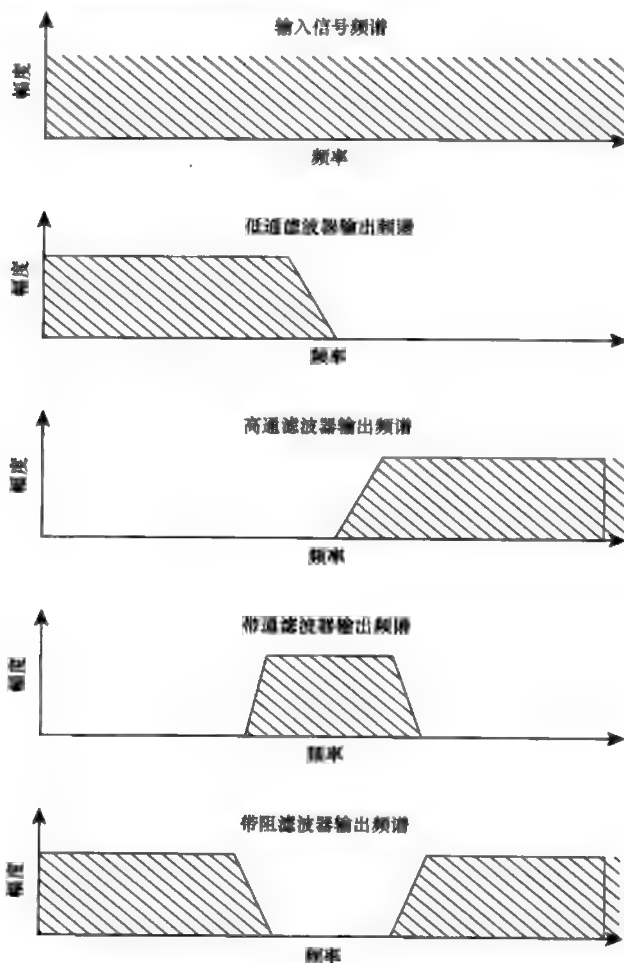


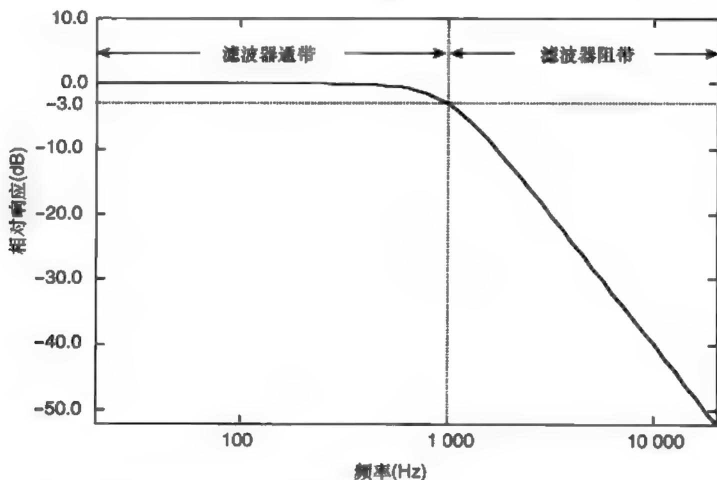
图 7.1 各种滤波器及其作用

使滤波对音色影响最小化的第一种方法是基于认识到许多乐器声并没有覆盖整个声频范围。几乎没有乐器其基频能够延伸到声频范围的最低端，而且许多乐器产生的谐波也不能到达声频范围的高频区。因此，从理论上说，如果滤波处理能够使乐器产生的频率成分顺利通过，就不会对听音产生影响。但实际上要做到这一点并不容易，主要原因有以下

2 个。

（1）滤波器形状：实际滤波器并不能从某个频率开始突然使信号不能通过，而是存在一个从顺利使信号通过到阻止信号通过的过渡过程，如图 7.2 所示。滤波器的截止频率一般定义为信号比通带信号衰减 3dB 时的频率，参看图 7.2。因此，如果滤波器以某一特定频率作为截止频率，那么在通带内就存在一个频率区域使信号的频率成分幅度受到一定程度的衰减，这个频率区域可以达到从截止频率开始的一个倍频程带宽。然而，当滤波器的阶次增加后，滤波器增益随频率衰减的斜率和截止频率处的陡峭程度都会增大。因此，在实际应用中，在分析相应乐器的频谱后，一般把滤波器的截止频率设置在乐器频率范围之外的某一频率。

图 7.2　实际滤波器特性



（2）时域特性的影响：滤波器也存在一个时域响应特性。任何一个减小信号带宽的滤波器都会延长信号的持续时间。对大多数滤波器电路而言，其截止频率附近的时域特性对信号的影响最大，而且截止频率处越陡峭，情况越坏。同时，就像滤波器形状的情况一样，这些影响会延伸到滤波器的通带里面。需要注意到，即使是陷波器，也存在陷波频带越窄，时域响应越长的效应。有趣的是，有些数字滤波器的时域响应特别差，以至在时域产生的人工声在输出端先于主信号输出。这些人为信号很难被掩蔽，因此对听音产生了影响。同样地，为了减小时域特性对声音的影响，滤波器的截止频率应设置在乐器频率范围之外的某一频率。

由于存在上述频域特性和时域特性的影响，要设计一个不会改变声音音色主观感觉的滤波器是十分困难的。

另一种使主观听音影响最小化的方法是基于认识到听觉是根据声音的频谱形状来判断音色的。因此，通过加强附近的频率成分，可以在一定程度上补偿由于滤除某些频率成分带来的影响。但是要注意到，这种方法产生的效果是有限的，处理也要适度。图 7.3 表示一个滤波器的频谱特性可以进行怎样的修正来补偿声音的频谱。图中截止频率附近的频谱被少量提升了 1dB ~ 2dB，用来补偿通带在截止频率附近的频率成分。

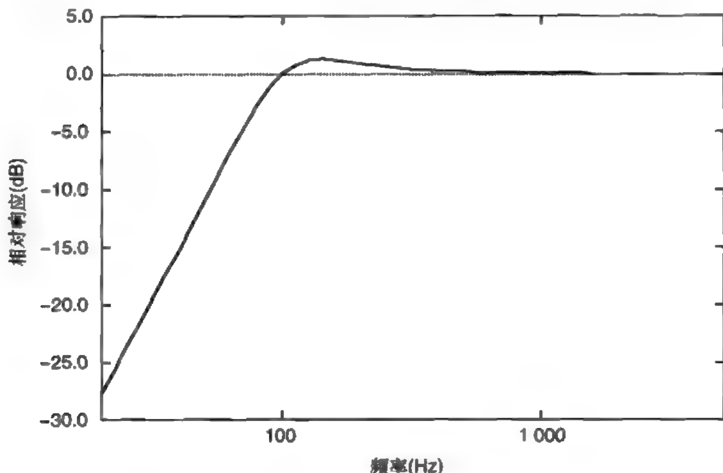


图 7.3 在通带边缘少量提升来补偿滤除的频率成分

7.2 均衡和音色控制

一种和滤波有关的非常重要的声音信号处理称为均衡。与滤波处理不同，均衡并不是要滤除某些频率成分，而是有选择地加强某些频率成分或减弱某些频率成分，以达到某种期望的声音效果。均衡处理可以看成是一个从原信号中加上或减去滤波器处理后的信号的过程，如图 7.4 所示。加上滤波器处理后的信号则使滤波器选择的频率得到加强，减去滤波器处理后的信号则使滤波器选择的频率减弱。滤波器可以是简单的高通或低通滤波器，实现对声音高频区和低频区的音色控制，也可以是钟形的带通滤波器。滤波器的截止频率可以是固定不变的，也可以是可变

的，取决于滤波器的设计要求。此外，带通滤波器的带宽一般可调，在较少的情况下，高通或低通滤波器截止频率处的衰减率也是可调的。滤波器参数可调的均衡器称为参数均衡器。然而，在实际应用中，特别是在调音台中使用的均衡器，为了经济和操作简便，一般使用较为简单的均衡器。在这类应用中，使用较多的是带通滤波器的截止频率可调，少数情况是高通或低通滤波器的截止频率也可调。还有另外一种均衡器，它是由固定通带、紧密相连的一系列带通滤波器构成，一般可以覆盖整个声频范围。这种均衡器称为图示均衡器，每个带通滤波器的典型带宽为 1/3 倍频程到 1 个倍频程。对于参数均衡器而言，其带宽可以控制得很窄。

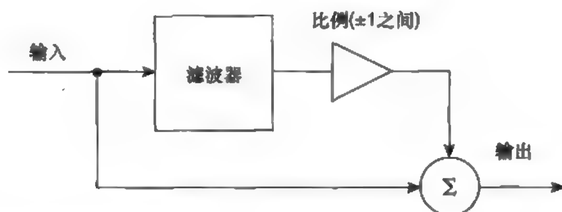


图 7.4　音色控制电路方框图

由于均衡器是由滤波器构成，因此对声音信号也存在上述时域特性的影响。这种时域特性影响对窄带均衡器尤为明显。不论使用增强还是衰减模式，相应滤波器产生的振铃声会持续相当一段时间。均衡器主要应用于 3 个方面，各自存在不同的声学和心理声学依据，将在 7.2.1 ~ 7.2.3 节中介绍。

7.2.1　校正录音过程中产生的不良频率响应

这个用途是早期经常使用的，现在，由于电声换能技术和电子技术的不断完善，一般不需要进行这样的处理。但是，有时因为声学环境和传声器摆位的原因，仍然需要做这样的频率响应修正。以下是针对这些问题需要使用均衡器的 3 种情况。

（1）用有指向性传声器进行近距离拾音：由于近讲效应，当有指向性传声器靠近声源拾音时，会使传声器的低频响应提升。这使得录制的声音低频变得很重，有些歌唱家经常有意利用这种效应来改善声音。这

种效应可以通过在传声器输出端使用低切滤波器去除，同时还能减小来自环境的低频噪声。值得注意的是，有些传声器设有内置均衡器，但是，一般来说，需要一个可变均衡器才能较好地补偿这种效应。

（2）补偿传声器的指向性频率特性：大多数实际使用的传声器不能保证在所有方向都有平坦的频率响应。一般来说，传声器指向性随频率增大而变强。为了拾取优良的直达声，大多数传声器设计有平坦的轴向频率特性，这就导致拾取混响声的频率响应在高频端有所下降。当传声器距离声源较近时，即以拾取直达声为主时，这种影响并不明显。但是，在进行古典音乐录音时，即当以拾取混响声为主时，这种效应对声音的影响就比较显著。可以通过在传声器输出端使用高频补偿来减小这种效应对声音的影响。

（3）补偿混响声场的频率特性：在许多演出场所，混响声场的频率特性并不平坦。因此，当拾取混响声占主导的声音时，会对声音的主观听感产生染色效应。一般混响频率特性的低频比理想情况有一定的提升，造成录制的声音低频较重。同样地，可以通过使用低切滤波器减小这种影响。

上述各种方法都是对录音时声学方面带来的频率特性缺陷进行补偿。要在这方面灵活掌握均衡器的使用，就必须了解在不同传声器摆位时，尤其是能够根据所拾取声音的直达声与混响声的比例，判断其可能对声音的频率特性产生的影响。

7.2.2 声源音色的修正

均衡器的主要作用之一是修正无论声学乐器还是电子乐器产生的声音音色，以满足艺术创作的需要。在这方面，均衡器可用于增强或减弱所选择频带的频谱，获得所期望的音色效果。例如，加强高频成分可以增加乐器声音的明亮度或活跃度，而加强低频成分则可以增加声音的“沉重感”或力度。均衡器只通过改变声音的频谱达到这些效果，而不改变音乐信号的包络或动态特性。任何音色的改变纯粹是因为用均衡器改变了音乐信号的长期平均频谱（long-term spectrum）。为了避免处理后的声音听起来不自然，这种修正也应该是有限度的。但是，在有些情况下，

某种奇异的声音效果又是人们想要的。

当使用均衡器修正乐声的音色时，值得注意的是要避免产生“听觉心理疲劳”现象。产生这种现象是因为听觉和大脑会适应声音，一段时间以后，会使音色改变的效应在听感上减弱，因此，往往导致操作者在听觉已经适应的频带进行更多的提升，或其他类似的调整。解决这个问题的唯一办法是，在听某段音乐时中间稍微间断一下，然后重新再听。请注意，这种现象可以发生在以正常声压级听音的时候，因此，与由于音量过大造成的暂时性听阈偏移的听觉疲劳现象有所不同。

7.2.3　混音时改善声音的平衡感

均衡器的另一个作用是在混音时调整声音的平衡感，尤其是可以使乐队中的某个声音“突出”或“隐蔽”。这是因为均衡器能够修正某一频带幅度，使某一特定频带的声音被其周围的声音或多或少地掩蔽。这就像第 4 章所述的歌唱家通过修正共振峰使歌声能够突出在伴奏乐队之上被听到一样。例如，某歌声的频谱完全被其他伴奏乐器的频谱所覆盖，如图 7.5 所示，显然其他乐器的频率成分把歌声的频率成分掩蔽了。通过适当降低乐队在 1.5kHz 附近的频率成分，同时对歌声在相应区域的频率成分进行提升，歌声的频谱就不会完全被掩蔽，如图 7.6 所示。这种处理在听觉上产生的效果是使歌声从伴奏声中凸显出来。同样地，如果在混音时做相反的处理，就会使歌声更不容易被听到。要成功地达到此类处理的目的，应使希望得到处理的频率成分处在均衡器加强或衰减的频带之内。因此，在进行这类处理时，应针对不同乐器选用不同的加强或衰减频带。同时还要注意提升或衰减要适量，以免声源的音色产生太大变化。

由此可见，均衡器在声音信号处理中具有广泛应用。然而，尽管它们具有很强的实用性，使用时还需要注意以下几点。其一，要避免使声音极度个性化，除非有意要这样做；其二，要避免不同均衡频带之间不必要的相互作用。作为一个简单的例子，试想想看如果对某个信号同时进行高频、低频和中频提升的结果会是怎样。由于均衡器不同频带频率响应之间的相互作用，结果是均衡后的信号频谱和原信号频谱完全一样，仅有的不同是信号的增益提高了。当对某个频段进行提升以后，发现这

个频段提升得有些过量，然后再去提升其他频段以对此进行补偿，这时就可能发生这种现象。

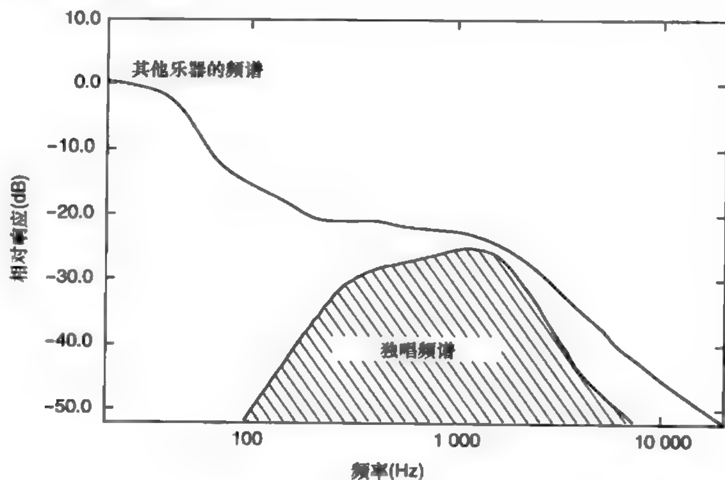


图 7.5　混音时被掩蔽的歌声频谱

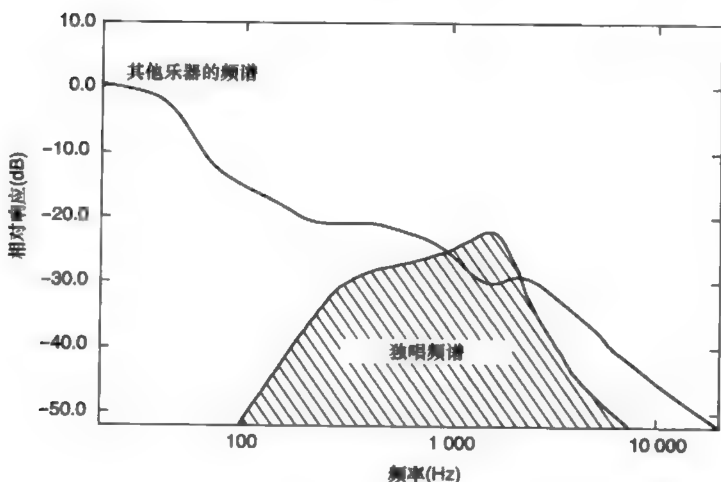


图 7.6　使用均衡器后使歌声不再被掩蔽的频谱

7.3　人工混响

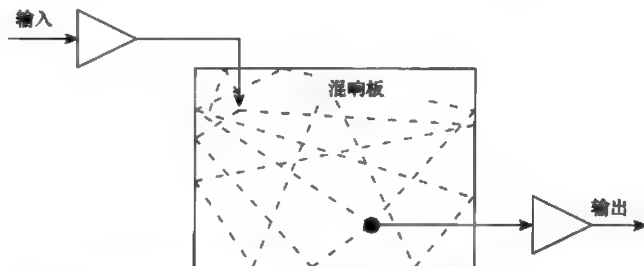
另一种常用的音乐信号处理形式称为人工混响。过去，人工混响主要通过使用电子机械装置如金属弹簧和金属板等实现，有时也使用

房间加人工混响。现在，人工混响主要通过数字信号处理技术实现，即首先将声频信号变换为数字信号，然后利用计算机运算处理实现人工混响效果。详细的数字人工混响处理技术不是本书讨论的重点，对此部分感兴趣的读者可参看参考文献 Zölzer(1997)。但是，数字人工混响的基本概念可以在不涉及详细内容的情况下进行介绍，由此足以认识到数字技术能使声音信号的存储更为方便，从而实现对信号的延时和精确控制。

7.3.1　模拟混响技术

这类电子机械装置是以需要加入混响的信号作为激励，以机械方式激发金属板振动，然后拾取其输出作为人工混响信号，如图 7.7 所示。由于金属材料的吸声很小，声波沿着金属板传播并来回反射，振动在完全消失之前可以持续相当长的一段时间，从而模拟出人工混响。然而，这种系统存在一些缺点，如所产生的混响听起来不太自然、对外界噪声和振动较敏感等。虽然最好的这种混响器能产生很好的混响声，多数这类混响器的声音听起来比较刺耳并带有金属感。尽管如此，从艺术的角度考虑，这类混响装置产生的声音过去是、现在还是被认为是一种有价值的选择。

图 7.7　板式混响装置



为了理解混响器背后的声学和心理声学依据，需要回顾在第 6 章提到的理想房间混响曲线，如图 7.8 所示。其按时间顺序主要包含 3 个重要方面：

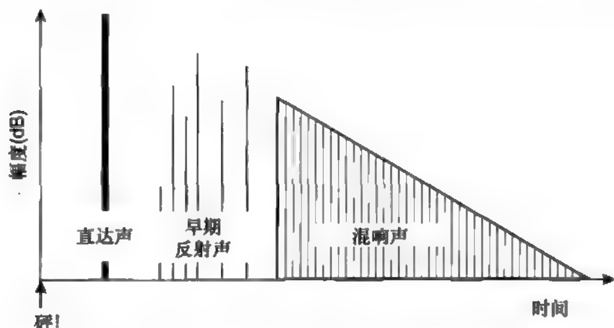
（1）初始延时；

（2）早期反射声；

（3）光滑流畅的、按指数规律衰减的、由密集扩散反射构成的混响拖尾。

从频率特性上看，混响拖尾的衰减在高频端应该稍微快一些，以模拟空气吸声的影响。它还不能有任何占主导地位的或可察觉的共振。

图 7.8 理想的混响特性



由金属制成的弹簧式或板式混响装置的声波传播速度与空气中的传播速度不同，其尺寸也远小于平均厅堂尺寸。设计者一般采用弹簧或板的最低速度模式激发振动，并使用某些材料来帮助达到此目的。这表明板或弹簧的有效声学尺寸很小，导致人工混响具有以下几个特点：

（1）实际上不存在初始延时，因为初始延时是由声波在材料中的传播速度以及进行发送和接受的换能器间距决定的，而这个间距非常有限。

（2）实际上不存在早期反射声。原因之一与上述相同，另一个原因是弹簧或板的边界数目有限，这使得声波快速转换为密集的混响声形态。同样由于边界较少，其反射声的密集程度也比实际房间小。

（3）共振频率的间隔变大，使混响声听起来比较粗糙，并带有振铃效应。

此外，由于金属弹簧或面板的吸声特性与空气不同，其吸声随频率增大而增大的现象不那么明显，因此，产生的混响声较自然混响明亮。常见的做法是用磁带录音机或数字延时器在调音台的“混响送出”和混响器的输入之间产生一个延时，来模拟真实混响效果。

7.3.2 数字混响技术

现代混响器主要利用数字信号处理技术产生模拟混响。数字技术克

服了前述电子机械混响装置的许多局限性。由于存储器价格较低，因此可以很容易地实现任意长度的延时，而设计的主要难点在于找到实现所需混响形态的算法。现代混响器的算法由 3 个不同部分组成，虽然它们可以在同一个硬件上实现，但算法完全不同。典型的算法结构方框图如图 7.9 所示，其中的 3 个方块分别完成以下工作：

（1）初始延时：如图 7.9 所示，一条简单的数字延时线用来模拟初始延时。一般初始延时时间可调，以满足模拟不同尺寸房间和不同效果的需要。例如，10ms ~ 20ms 的延时使直达声和混响声之间有一个适当的间隔，而 50ms 的初始延时则会明显感觉到回声存在。

（2）早期反射声模拟：一条带有抽头和乘法器的延时线用来模拟早期反射声，如图 7.9 所示。抽头的位置和相应的乘法因子都是可调的。抽头的位置决定早期反射声的延迟时间，而乘法因子决定早期反射声幅度大小。由于乘法和抽头都需要运算时间，因此能够进行这样计算的数目是有限的，主要取决于可用的信号处理容量。一些较复杂的混响器往往使用更多的乘法器和抽头，甚至通过使用滤波器来实现早期反射声随频率变化的功能。

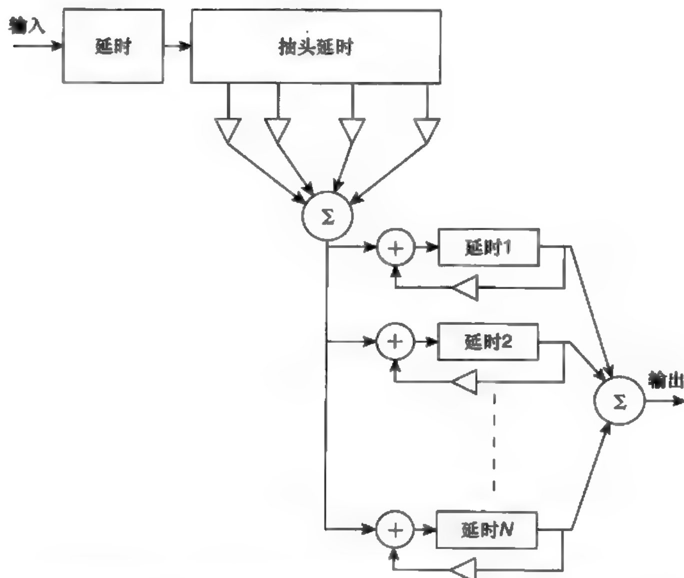


图 7.9　数字混响器工作原理方框图

（3）混响拖尾模拟：这部分使用了多条不同长度的延时线。与早期

反射声模拟的不同之处在于，每条延时线的输出都被反馈到输入端，从而形成了循环延时线。反馈通路中的乘法因子必须小于 1，以保证声能按指数规律衰减。乘法因子越接近于 1，衰减时间越长；反之，乘法因子越小，混响拖尾就越短。如果乘法器用一个滤波器代替，其输出随频率变化，假设其频率响应是高频较低频有所降低，那么其产生的混响在高频衰减得比低频快。这个技术可用来模拟空气吸声对混响的影响。“明亮的”和“暗淡的”房间模拟的主要区别在于前者比后者需要更长的高频混响衰减时间，而后者在反馈回路中需要更多的高频衰减。

混响模拟的主要困难在于必须有足够多的不同长度的延时线，才能产生足够密集的反射声，保证声能平稳地衰减。这是因为每条延时线只能模拟出一个房间内循环传播的声音路径，实际声场却有许多这样的路径。延时线的多少也会影响混响频率特性，因为每条循环信号路径代表房间的一个模式结构，正如前面第 6 章所述。循环延时线的频率响应如图 7.10 所示，它包含一系列等间隔的共振峰，其频率间隔与延时线的长度成反比。共振峰的带宽由反馈量的大小决定，当反馈量接近 1 时，带宽变得很窄。事实上，每条循环延时线都能有效模拟出一组共振模式（见第 6 章），其座落在模式频率坐标系中的一条径线上，如图 7.11 所示。由图可知，要模拟出房间内所有可能的路径是非常困难的，因为每一条径线都需要一个独立的共振模式。因此，在这方面，数字混响器遇到和前述

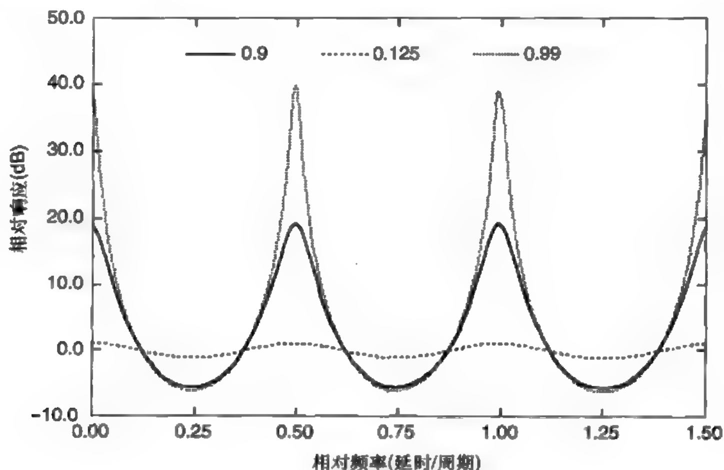
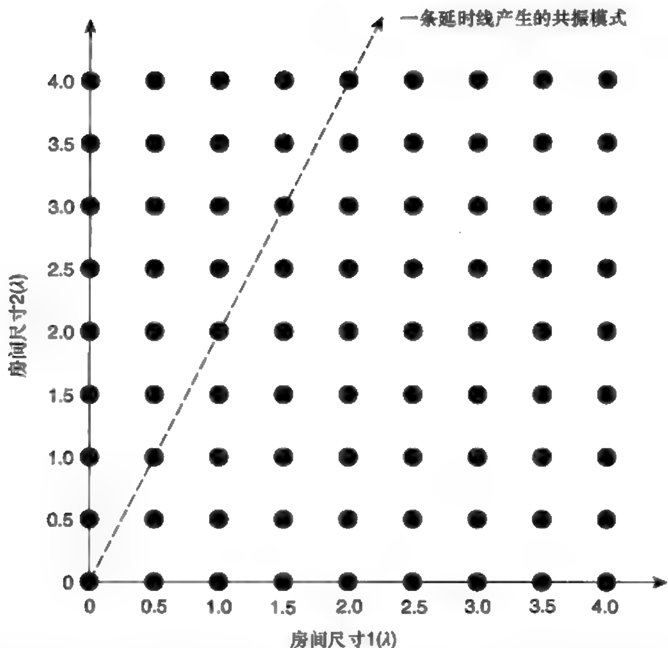


图 7.10　循环延时线的频率响应（不同曲线对应不同反馈量）

图 7.11　延时线与共振模式的对应关系



的板式或弹簧式混响器相同的问题，即不能有效模拟出实际房间反射声和共振模式的密度。很显然，混响器延时线的多少与可用的计算容量有关，因此就与造价有关。这是价格较低的混响器必须做出妥协的一个方面，以达到降低成本的目的。然而，还有许多不同方法可以改善反射声的密度，但是这些方法大多数是有专利的，它们包括各个循环延时线之间的互耦合技术、使乘法因子随时间连续变化等。

到目前为止介绍的混响器是单声道混响器，其输入信号和输出信号都是单声道的。图 7.12 所示为一些立体声混响器的工作原理方框图。图 7.12(a) 为一种由单声道信号产生立体声混响的方法，即将单声道信号分别送到两个不同的上述类型的混响器中，由它们产生所需的左声道和右声道混响信号。最重要的一点是两个混响单元不能使用同一组混响系数，否则将产生两个完全相同的混响输出。因此，这两个混响器需要满足以下的相同点和不同点：

(1) 两个混响器的初始延时和早期反射声形态必须不同，因为对真

实房间里的某个声源而言，不同位置其初始延时和早期反射声形态是不同的。

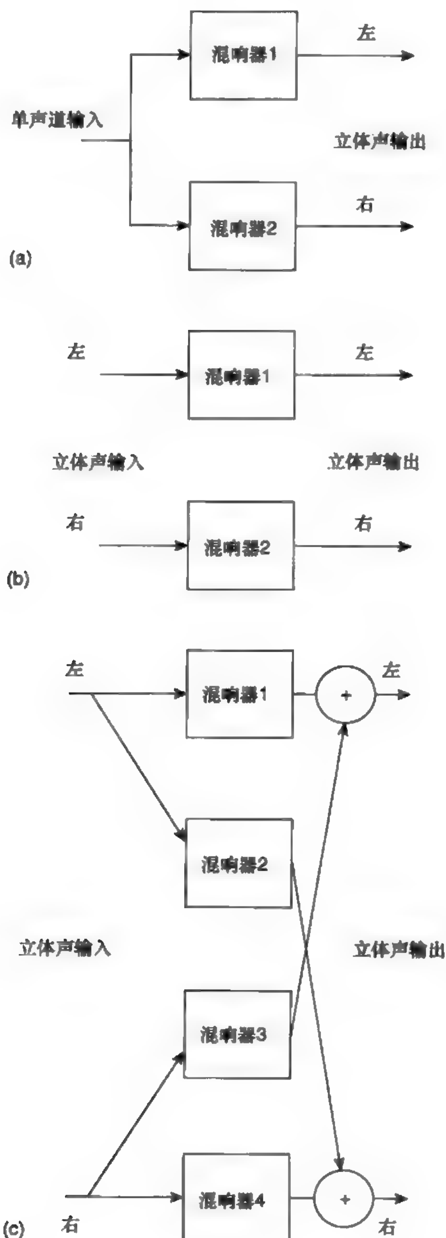


图 7.12 （a）单声道输入立体声混响器；（b）双声道输入立体声混响器；（c）改善的双声道输入立体声混响器

（2）两个混响器的混响拖尾必须具有相同的衰减速率，但是，两个混响拖尾必须进行去相关处理，因为在理想扩散声场中，来自不同方向的声音之间是非相关的。

还需要注意的是，从原理上说，当声像调整到声音舞台的不同位置时，早期反射声和初始延时也要随之变化，实际房间的情况就是如此。一个较简单的解决方案是只改变两个混响器的输入信号，但是，这样做会产生不自然的听音感觉，因为当声像位置改变时，虽然使早期反射声的幅度发生了变化，但同时也使混响拖尾的幅度发生变化，而这与实际声场情况不符。极端的情况是当声像完全调整到左边或右边时，其中一个混响器的输出可能为零。而实际房间声场的情况并不是这样的，正如第 6 章所述，混响拖尾不应该随声源位置的变化而变化。

图 7.12（a）所示的立体声混响器可用于单声源的情况，而不适用于包含多个声源并且每个声源必须调整到声像舞台的不同位置的情况。一种可能而且简单的解决办法是采用图 7.12（b）所示的使用两个独立的混响器。这两个混响器和前述的单声道到双声道的混响器具有相同的特性。但是，这种方法也存在混响拖尾随声像位置变化的缺点。一种更好的方法如图 7.12（c）所示。这种方法增加了另外两个混响器，用于在两个声道混响声的基础上，增加交叉馈送的混响声。这两个额外混响器的初始延时、早期反射声形态和混响拖尾的要求与前述的单声道到双声道的混响器相同。这种方法的优点是，就像在实际声场中一样，混响拖尾的幅度不会随声像位置的变化而变化。这是因为当声像改变位置时，互耦合通道就会把必要的混响声送到另一个通道，使两个声道的混响声幅度保持不变。此外，由于调整声像位置实际上是改变左、右声道信号的相对幅度，因此，当声源在声像舞台上改变位置时，其早期反射声形态、相对幅度、初始延时随之变化。虽然早期反射声形态的改变并不完全和实际声场一样，例如，早期反射声的时间位置的改变并不和实际声场完全相同，但是，它却为更真实地模拟实际混响声提供了必要的条件。图 7.12（c）所示的方法也可以用于从单声道产生立体声混响的方法中，这样，当声像位置改变时，混响器能更容易地改变早期反射声形态。

7.4 合唱、加倍（ADT）、梳状滤波和镶边效果

合唱、加倍（Automatic double tracking，缩写为 ADT）、梳状滤波和镶边效果采用与人工混响相似的技术和算法。加倍、梳状滤波和镶边效果都需要使用延时线。

（1）ADT：这项技术是利用一条延时线模拟两个相同的声音几乎同时在演唱的效果。典型的信号处理方法如图 7.13 所示。延时必须足够大，使得延时声能作为单独事件被听到。一般延时设在大于约 30ms。但是，即使这样延时声也很难作为第二个声音被听到，因此实际使用时还需要对信号进行延时调制、音调偏移等处理，使加倍效果更好。

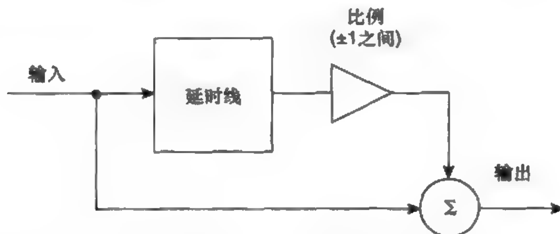


图 7.13 加倍（ADT）信号处理方框图

（2）梳状滤波：这种效果利用了将延时线的输出加入到未延时的信号所产生的梳状滤波效应，如图 7.14（a）所示，其产生的等效频率响应如图 7.14（b）所示。通过对延时线的延时进行低频调制，可以产生愉快的声音“旋动”效果。这种效果是滤波器的谷点在声音所覆盖的整个频率范围内移动的结果。有趣的是，如果让谷点位置保持静止不变，则谷点的作用很难被听觉察觉。

（3）镶边：这个效果和梳状滤波相似，所不同的是它使用了一条循环延时线，而不是一条简单的延时线，如图 7.15（a）所示，这使得频率响应产生一系列等间隔的共振峰，如图 7.15（b）所示。同样地，延时时间被调制，使共振峰位置不断变化，从而产生随时间变化的听音效果。就像梳状滤波效果一样，这种共振峰位置随时间的变化加强了镶边效果。

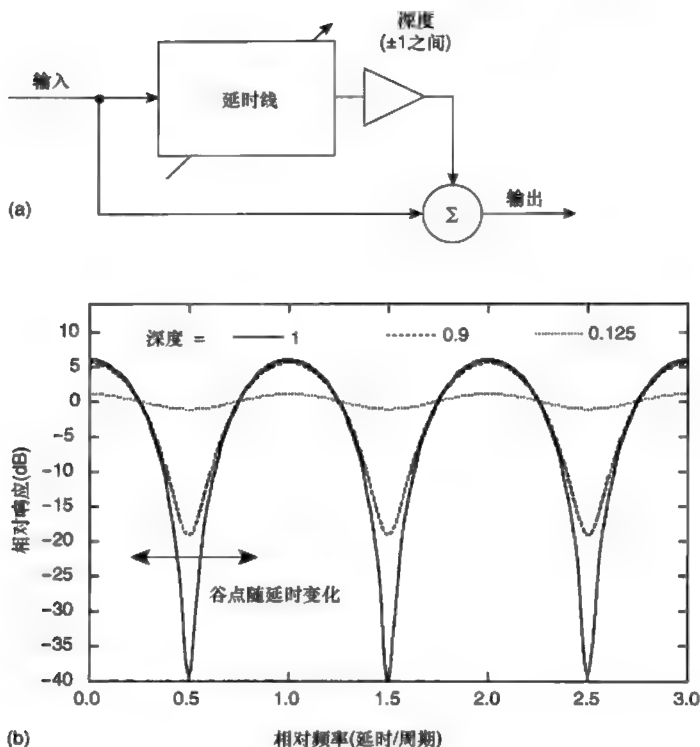
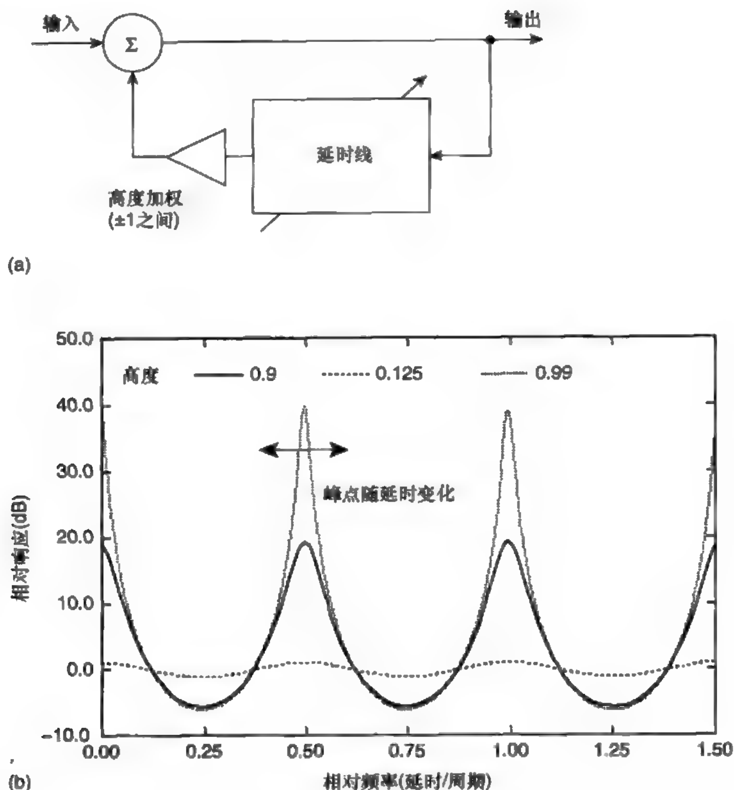


图 7.14 梳状滤波效果
(a) 信号处理方框图；
(b) 频率响应

这些效果处理开始于磁带技术时期，当时是通过同时播放两个相同的声音拷贝来达到这种效果。事实上，“镶边”这个词来自当时产生效果波动的方法，即通过物理上操作重放声音的磁带录音机的磁带凸缘，略微地改变两个磁带之间的延时。

另一个有关这些效果的有趣现象是，当声源或处理电路中存在随时间变化的因素时，产生的效果会好得多。这是因为听觉对变化的声音非常敏感。这种现象也反映在混响器中。如果混响器输入的声音在频率和幅度都是恒定的，那么很短时间以后混响效果就很难听出来，这是因为虽然混响器的输出发生了变化，但它是静态的，因此很难被听觉感受到。如果输入声音信号有颤音，那么混响效果就容易察觉得多，因为颤音意味着音高随时间变化。也就是说，由于输入信号的音高随时间变化，因此听音者在任何时刻听到的混响声都不同。这也是颤音在音乐演奏、声音合成以及实际演唱中占重要地位的原因之一。

图 7.15　镶边效果
（a）信号处理方框图；
（b）频率响应

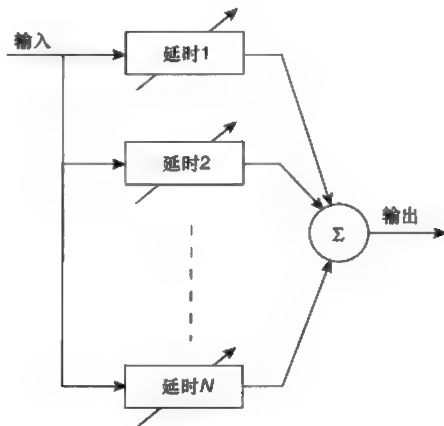


（4）合唱：合唱效果试图从单个声源获得一组或合唱的听音感受。这种声音效果在 20 世纪 70 年代的各种合成器中非常流行，目的是产生交响乐队中弦乐组或铜管组的合奏效果。为了达到这一目的，合唱效果器必须能够模拟出同一个声源的不同版本，而仅仅通过延时来复制它们是不够的。实际上，在同一种乐器或歌声的合奏中，每个乐器都有轻微的音高和音高变化的差异，同时也有轻微的音色差异和幅频特性差异，最后还存在每个音符起始时间的差异。

一种获得合唱效果的方法如图 7.16 所示。图中声音信号被同时馈送给多条延时线，各延时线的延时是可变的。延时可以以低频噪声或低频正弦波进行调制，但最重要的是每条延时线的调制频率、延时或噪声源必须不同。改变延时线延时的作用是改变输出的频率，因为输出读出的速度会与输入略有不同。因此，采用多条独立调制延时的延

时线可以使各个输出在频率和时间上有所不同，从而产生多个乐器合奏的效果。

图 7.16　合唱效果信号处理方框图



还有其他方法也可以获得合唱效果，例如利用时变共振滤波器、音高偏移等。尽管方法不同，但它们的共同点都是试图通过某种方式在一定程度减小复制信号和原信号之间的相关性。

到目前为止我们完成了对包含延时和混响的信号处理的讨论。在一些应用中，提到了采用音高偏移的信号处理技术来加强混响和延时效果。这个信号处理技术是下一节要讨论的主题。

7.5　音高处理和时间修正

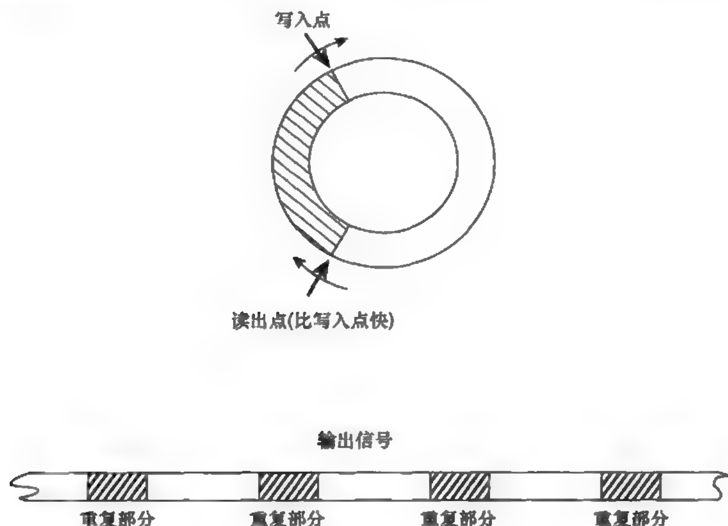
音高处理是用于描述针对音高的信号处理的一般性词汇，后面我们还将看到，时间修正是一种与之相关的信号处理。在现代音乐制作中，尤其在再混合和俱乐部风格的音乐中，音高和时间修正是这个行业的基本处理手段。大多数音高处理包括以某种方式修正信号当前的音高，例如简单地将音高向上或向下进行少量的偏移、施加一个额外的音高调制包络或者甚至从单个声源产生和声。在许多情况下，音高修正可能还包括信号的一些相关时间修正，也可能是要除去声源的时间修正效果。请注意采样式合成器从基本原理上看就是一种音高处理器，因为它是从一个或者很少几个波形蓝本中提取大量不同的音高。在后面我们会发现，

某些采样式合成器面临的问题和信号音高修正处理中遇到的问题是完全相同的。

修正音高的最简单方法是将需要偏移的信号存储在一个环形缓冲寄存器中，称为环形是因为数据被打包成环形，当到达存储器的末端时，形成了一个圆。然后数据以较高或较低的速度被读出，取决于要将音高向上或向下偏移。

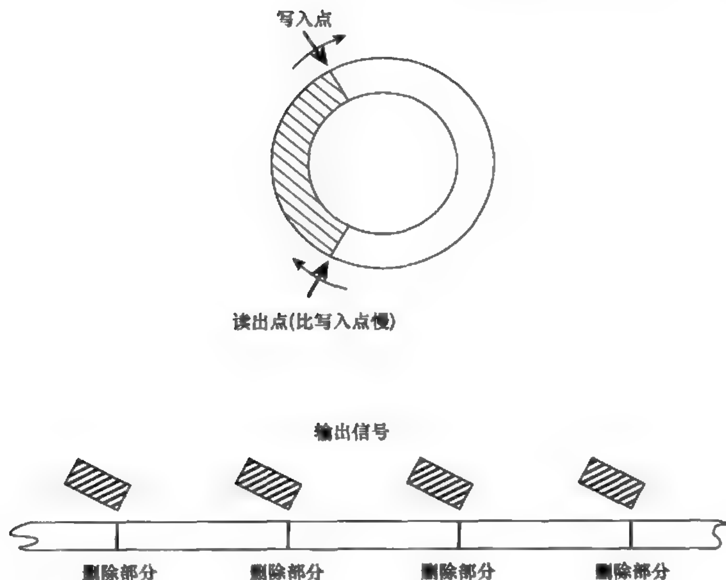
当需要将音高向上偏移时，读出的速度大于写入的速度，并且读出点会赶上并超过写入点。由于缓存器是环形的，储存在写入点前面的数据就是之前刚刚被读出的数据。因此，读出点赶上并超过写入点的结果是重复读出一部分信号，如图 7.17 所示。重复的数据量与缓存器的长度有关，重复发生的频率取决于所需的频率偏移量；重复频率越高，则频率偏移越大。

图7.17 向上的音高偏移



当需要将音高向下偏移时，读出的速度比写入的速度慢，写入点会赶上并超过读出点。由于缓存器是环形的，读出点前面的数据正好是之前写入的数据，因此，写入点赶上并超过读出点的结果是删除了一部分信号，如图 7.18 所示。删除的数据量与缓存器的长度有关，删除发生的频率取决于所需的频率偏移量；删除数据量越大，则频率偏移越大。

图7.18　向下的音高偏移



这项技术虽然简单，但也存在以下几个问题：

（1）波形的不连续性：信号的重复或删除都会使波形不连续。由于重复和删除是以恒定速率出现，速率又是由所需的频率偏移量决定，因此对输出声音的影响是以相同的频率出现嗡声。有多种方法可以减小这种效应。一种方法是使用大缓存器使重复或删除的数据长度尽量长，这样在频率偏移量不变的情况下会降低单位时间重复或删除数据的次数。但是，延时和重复或删除效应的可察觉性都对缓存器的长度有所限制。从掩蔽效应考虑，缓存器允许的最大容量约为 30ms。此外，还可以采用一些算法使波形的不连续性最小化，例如在连接处逐步减小而不是突然切断、只在同一方向重复或删除波形中穿过零的部分、重复或删除的信号长度与音高的周期相关等。这些技术可以单独使用，但是更经常与其他技术结合使用。在这些技术中，只有与音高周期相关的重复或删除技术不会使听觉察觉，因为信号处理所产生的额外谐波成分是与音高相关的，即所产生的频率成分属于原有的谐波成分，因此很容易被本身固有的谐波成分掩蔽掉。在其他处理技术中，所产生的谐波成分与音高无关，因此较容易被听觉察觉。平滑转换和利用波形中同一方向的零交叉的目的是减小附加频率成分的幅度和频率范围，从而增加被波形本身的频率

成分掩蔽的可能性。遗憾的是，与音高同步的重复或删除技术只对非复音音乐（每次只出现一个音符）的声源可行，这是因为要找到重复或删除的长度正好与和声中所有音高成整数倍关系，特别是当采用平均律时，是极为困难的。因此，实际音高偏移器为了满足各种声源的需要，必须采用其他处理技术，这也有效避免了音高检测方面可能产生的任何差错。

（2）共振峰的偏移：这些算法都是通过改变输出数据的速率来达到音高偏移的目的，这等效于提高或降低磁带录音机的转速。重复或删除部分数据的作用是使录音机能以正常速度重放声音而获得音高偏移的效果。因此，就像磁带录音机改变音高那样，采用这些算法的效果是使声音的频谱按所需偏移量向上或向下偏移，如图 7.19 所示。在个别情况下，这样的处理会产生“花鼠”和其他效应。请注意在采样式合成器中也会产生这种效应，原因是相同的，它们也是通过改变采样速率来改变采样的音高。由于这种效应的存在，为了避免输出的声音听起来有异样的感觉，必须对音高偏移量有所限制。尽管如此，音高偏移技术对产生各种声音效果是非常有用的，如合唱效果、ADT 效果、使音质丰厚等，甚至可以用于修正已录制节目的微小的音调误差。在采样器中，这种音高偏移的限制可以通过在整个乐器音高范围内使用多种采样得到缓解，从而限制任一给定采样的音高偏移范围。遗憾的是，这不适用于已录制的或现场声源。然而，音高偏移技术的一个主要应用是修正录制的音乐以不同的速度重放所带来的不良效果，在广播中心经常需要进行这样的处理。它可以控制节目播出时间的长度，使其正好符合所预留的时间间隔。这项技术也广泛应用于舞蹈音乐和再混合音乐行业，它允许 DJ 或制作人去配合不同类型音乐的节奏，除了找准音调外，有时还需要做一些音高偏移的处理。在这类应用中，这种音高偏移算法的共振峰偏移特性正好满足了需要，因此，这项技术在需要进行时间修正的音乐制作行业得到了广泛应用。

还有其他对声音信号音高进行修正的方法，而且不会像前述较简单的方法那样引起共振峰的偏移。其中一种方法是按照第 4 章介绍的声学模型将声音按声源和声音修饰滤波器进行分离，然后将音高偏移算法仅应用于声源部分，最后将音高偏移后的声源部分再与滤波器特性相结合，得到音高偏移但共振峰没有改变的输出。这是因为共振峰特性是由滤波

器决定的，而滤波器的各频率成分并没有进行音高偏移处理。滤波器特性的剥离可以通过将信号经过自适应反转滤波器（译者注：其传输函数为原滤波器的倒数）处理后实现，整个偏移处理方框图如图 7.20 所示。由于需要大量的计算来实现反转滤波器，因此这种方法在过去是无法实现的。但是，对现在而言，实现这样的信号处理是较容易的事情。

图 7.19 音高偏移对信号共振峰的影响

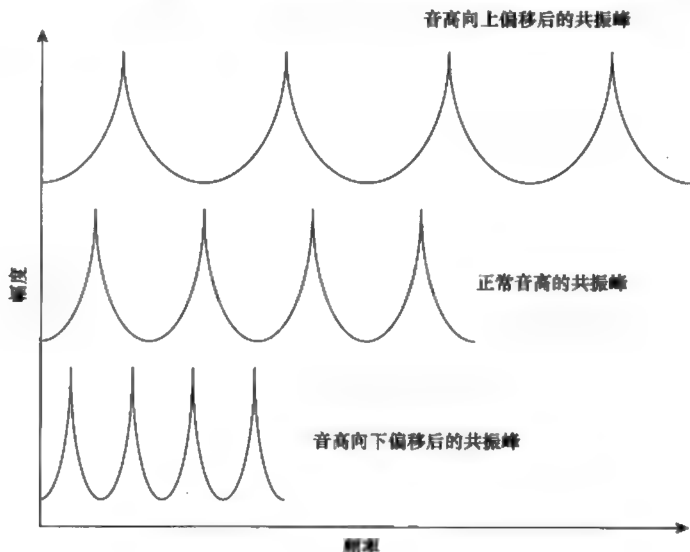
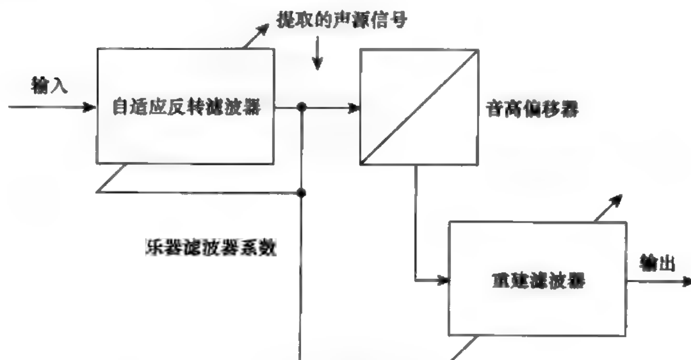


图 7.20 采用自适应反转滤波器的音高偏移方法



另外还有一种不影响共振峰的音高偏移方法也是基于乐器的“声源—滤波器”声学模型。这种方法的出发点是，正如图 7.20 所示，唯一能使音高改变的是输入到共振峰滤波器的声源信号的速率。而且，当声源信

号是一个脉冲信号时，滤波器函数就是简单的输出波形的读出。这些波形值对应于实际输入波形在一个音高周期的值，如图 7.21 所示。因此，音高偏移算法实际上是输入信号音高（或周期）的检测，随后是最近的一个音高周期的波形提取。这样，每当音高检测产生一个激励脉冲波时，这个波形就被同时读出。信号处理过程如图 7.22 所示，所产生的输出波形如图 7.23 所示，分别对应于向上和向下偏移的情况。

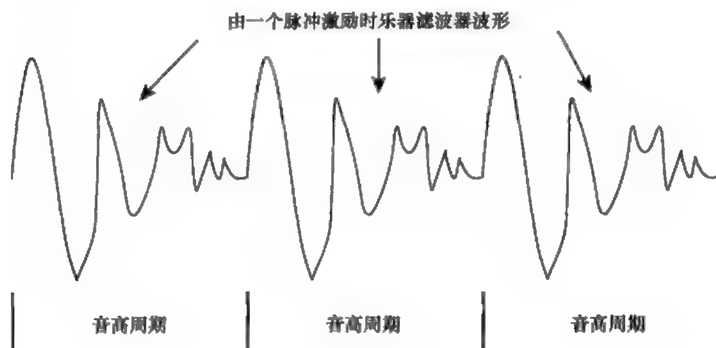


图 7.21 某乐器的理想化波形

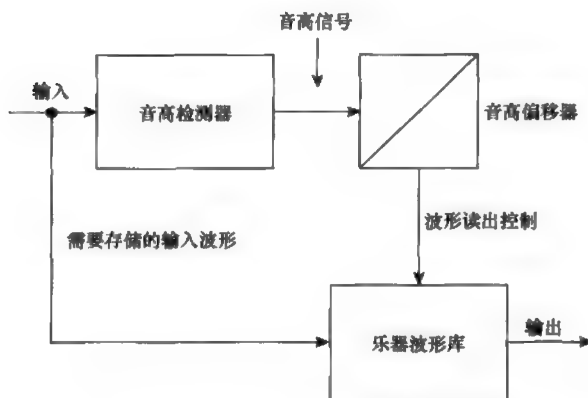
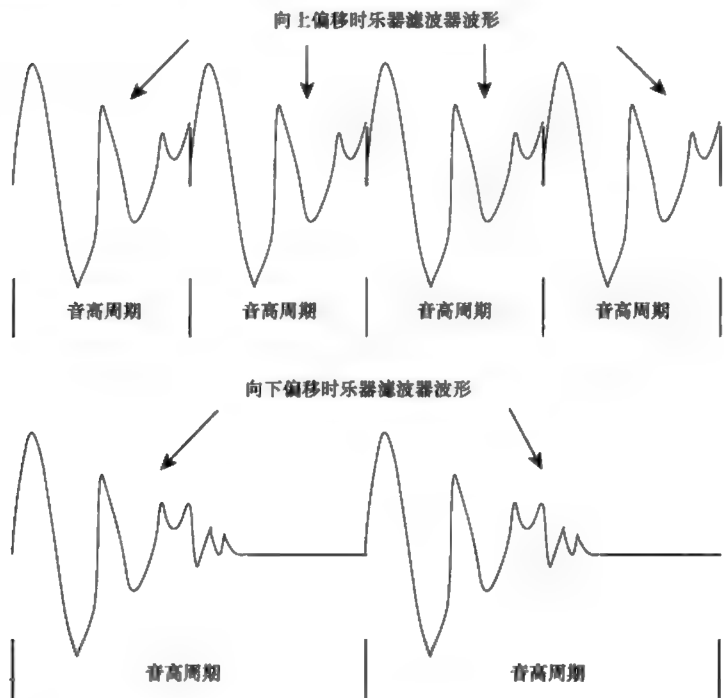


图 7.22 基于波形的音高偏移器信号处理方框图

需要注意的是，能够提取的波形的多少与音高有关，一般来说，仅仅是用脉冲激发的乐器共振的完整波形（脉冲响应）的一部分。由于在实际操作中不可避免的滤波器脉冲响应截断，偏移处理后的输出会产生振铃效应。当截取的波形用于音高向下偏移处理时，在每个音高周期会出现一段空隙，如图 7.23 所示。这会使处理后的声音产生蜂音，因为这样的处理会使共振峰明显加宽。当音高向上偏移时，每个音高周期的波形会

部分重叠。一种简单的处理办法是当下一个周期到来时就停止读出波形，这样可以提高计算效率，但是会进一步强化铃声效应。另一种处理办法是将重叠部分相加，这样会增加额外的计算量，从而限制了最大向上偏移量。例如，如果最多能够将 4 个采样值相加，那么音高向上偏移的最大量可达 4 倍或两个倍频程。由于需要进行音高检测，这项技术只适合于非复音音乐声源，原因如前所述。然而，尽管存在上述限制，这项技术已被证明是实现音高偏移的有效技术。即使音高偏移量较大，其输出声音的质量仍然是可接受的，因此，它可能成为另一个有效的采样式合成器技术的基础。

图 7.23　基于波形的音高偏移器的输出波形

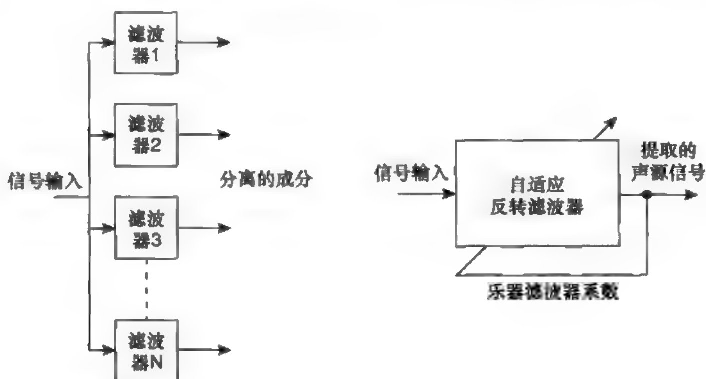


7.6　声音的变形与合成技术

其他既利用“声源—滤波器”声学模型又利用人耳对声音的感知特性的声音处理还有声音的变形与合成技术（sound morphing and

vocoding）。在这类信号处理中，输入声音信号首先被分离为声源及其不同的频率成分。这可以通过自适应反转滤波器和一系列并联滤波器来实现，如图 7.24 所示。理想情况下，滤波器的带宽应和临界频带相近，大约为 1/3 倍频程带宽。然而，在实际应用中，为了减少所需的滤波器数量，选取的滤波器带宽会较宽一些。一旦声音被分离出声源和各频率成分后，就可以进行一些有趣的信号处理。例如，从不同声音提取出来的声源信号可用作乐器滤波器函数的输入，从而得到二者结合后的产物。因此，就可能产生出海滨的低语和钢琴音色的语声。这种声音信号处理方式称为声音合成技术（vocoding）。另一种类似的处理方式是将不同的声源和不同的滤波器函数相结合，产生通常听不到的一些声音，称为声音的变形技术（morphing）。例如，女高音歌手的声源与高男中音（counter tenor）的滤波器函数结合后，总体上可以得到一个类似于做阉割手术歌手的歌声，而在现代社会这样的声音已经不存在了。声音变形技术也可以进行动态信号处理，实现两种不同声音之间的顺滑转换，例如，将蜜蜂的“吱吱”声变换到一群蜜蜂的声音。

图 7.24 提取声源和频率成分的信号处理方框图



声音变形处理需要将不同声音的声源分离出来，并分解成不同的频率成分，然后通过某种方式重新组合，图 7.25 所示为典型信号处理方框图。声音合成与声音变形使用的硬件和算法非常相似，所不同的是声音合成一般使用整个的其他声源作为激励信号，而不需要将它们分解成不同的频率成分。通过使用整个的其他声源信号，可以产生例如风在低语的听音幻觉，如图 7.26 所示。而另一方面，声音变形处理则需要通过更多地对单个频率成分的控制来实现它的声音效果。

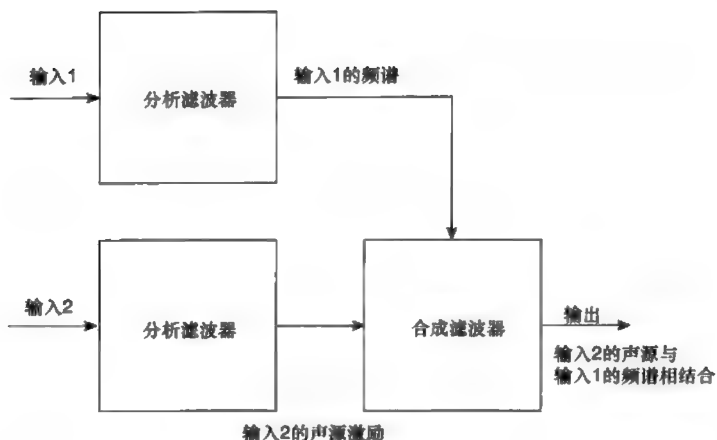


图 7.25　声音合成器信号处理方框图

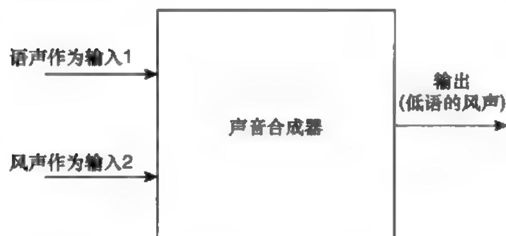


图 7.26　用合成器将声音频谱作用于另一个声源

这两种效果器都依赖于"声源—滤波器"模型实现信号处理，因此一般对单音或非复音音乐声源的处理比较有效。但声音合成技术也能轻易地用复音音乐信号作为可替代的声源。为了充分利用人耳的听觉特性，两种信号处理方法都倾向于使用带宽接近于临界频带的一组滤波器。

7.7　空间感处理

空间感处理是现代音乐制作的一个重要组成部分，有多种不同的系统目前正在使用中（参考本书 CD 第 78 条）。

7.7.1　立体声

最简单的能够营造声音空间感的系统是双声道立体声重放系统，它

主要是通过强度差信息给听音者产生声音来自两个扬声器之间某个方向的感觉。这主要基于两个事实：一是当频率高于 1kHz 时，由于人头对声波的遮蔽效应，听觉主要依据两耳声音的幅度差判定声源的方向；二是在低频时，双耳信号之间的延时也会产生一定的声级差。因此，强度差立体声可以营造一系列令人信服的声像，正如 2.6.6 节所述。

7.7.2 5.1 环绕声

5.1 环绕声系统是在立体声基础上发展起来的空间感处理系统，主要应用于电影和视频图像的播放。5.1 环绕声系统不仅能够提供清晰的对话，而且能够提供立体声音乐、效果声和听音环境感，典型的扬声器布置如图 7.27 所示。图中，除了传统的两个前置扬声器外，还增加了下述的其他扬声器，以满足更高的声音重放要求：

（1）中置对话扬声器：影片对话内容一般由中置扬声器重放，因为实践证明这样可以获得比两声道重放更好的语言清晰度。有趣的是，听觉并不能察觉到语言重放是单声道的。这是因为在声音定位时，视觉起较为重要的作用，因此，听觉总能感到声音来自屏幕上说话人的方向。

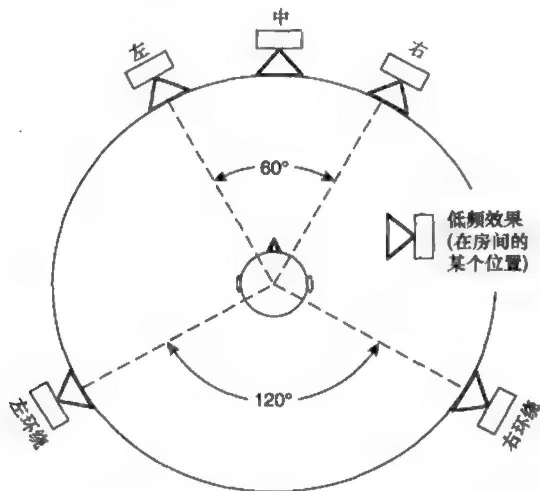


图 7.27 典型 5.1 环绕声扬声器摆位

（2）环绕声扬声器：环境声和效果声是由后置的环绕声扬声器提供的。由于它们不是用来产生声音的方位感，因此馈送给环绕声扬声器的

声音的相关性应尽量小，与前置扬声器信号的相关性也要尽量小，从而使听觉产生声音不是来自特定方向而是来自四面八方的感觉。

（3）低频效果扬声器：在电影和视频节目中，经常会出现如爆炸声、撞击声等声音效果，这些声音包含丰富的低频成分甚至是次声波成分，因此，增加一只低频扬声器是非常必要的。一个专门用于重放低频的扬声器可以满足重放这些效果声的需要。值得注意的是，尽管低音效果扬声器也存在于许多重放音乐的环绕声系统中，但是，它们从来不是用来重放音乐信号的。

近年来，在市场上还出现了六声道或更多声道的系统，有关这方面的更多信息可参看参考文献 Rumsey(2001)。

7.7.3　声场式立体声

另一种具有空间感的声音重放系统称为声场式立体声（Ambisonics）。这是一种三维的声音重放系统，其出发点是模拟出房间内某一点的声场。为了模拟某一点的声场，必须得到以下几方面信息：（1）该点的声压；（2）从左到右方向的振速；（3）从前到后方向的振速；（4）从下到上方向的振速。这些分量可以通过使用 1 只无指向性的压强式传声器和 3 只八字形指向性的压差式传声器测量得到。3 只压差式传声器主轴方向互成正交关系，如图 7.28 所示。这样拾取的 4 个声道信息代表录音点的一阶声场。这一点的声场可以通过扬声器阵列在房间里重现。至少需要 8 只重放扬声器，典型摆位如图 7.29 所示。这种系统和立体声重放不同，没有特定的扬声器被摆放在特定的方向，如左、右扬声器等，而是所有扬声器一起工作，在房间里产生所需的声场。也就是说，所有的扬声器共同工作来产生声压分量，并通过馈送不同扬声器合适的幅度和相位关系的信号来产生所需的振速分量。例如，前后分量可以通过馈给扬声器 1 和 2 相同强度的前后信号、馈给扬声器 3 和 4 相同强度但相位相反的前后信号、不馈送信号给扬声器 5 到 8 来合成。之所以不馈送任何信号给扬声器 5 到 8，是因为它们正好处在前后八字形指向性的零点位置。这种信号分配方法只对图 7.29 所示的扬声器摆位适用。如果扬声器位置有所改变，那么信号馈送的加权方式也要改变，而且加权方式与拾取不同声

道的传声器指向性有关。该系统可以通过减少声道数得到简化。当减少一个声道时，系统就简化为二维声场重放系统，如果声道数减少到两个，就简化为双声道立体声系统。从原理上说，还可以在拾取振速分量的基础上，增加对加速度分量的拾取，从而提高再现声场的精度，但实现起来比较困难。

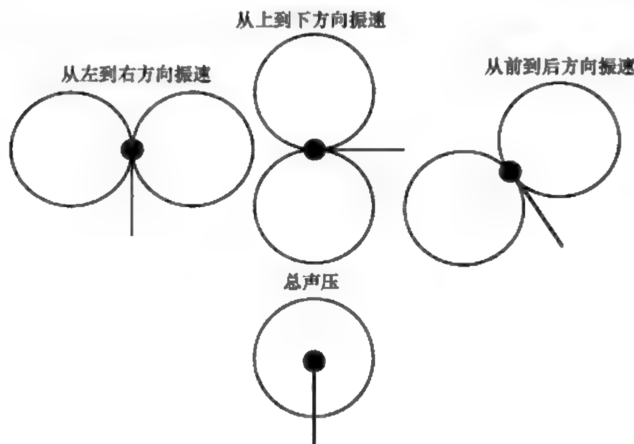


图 7.28 声场式系统拾音信号（注意：所有传声器置于同一点）

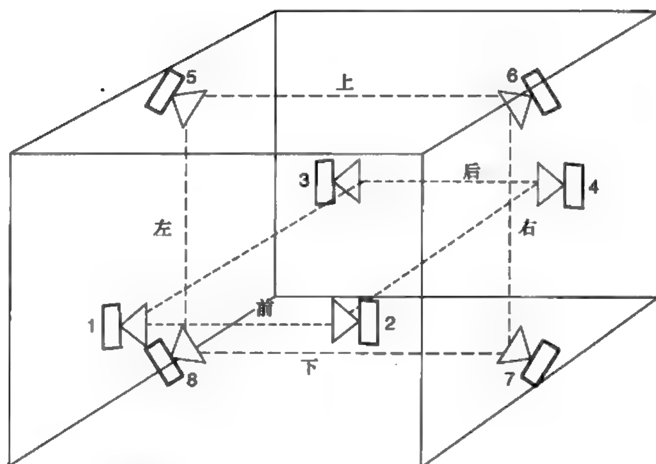


图 7.29 声场式系统的扬声器摆位

7.7.4 仿真头立体声

正如 2.6.6 节所述，当录制的声音需要用耳机重放时，采用仿真头系

统更为合适。仿真头录音的重放效果非常逼真，特别是当仿真头录音采用的是听音者本人的双耳模型时，重放效果更佳。但是，当采用的是其他人的双耳模型时，重放效果就没有那么逼真了，因为每个人只熟悉自己的头部和双耳的频率响应。实验表明，选用一个较为简单的人头模型似乎最为合适，这种仿真头只模拟出人头遮蔽效应和双耳时间差效应，那些容易使听音者迷惑的方向信息反而减少了，因此重放效果更好。仿真头立体声的另一个问题是，当听音者转动头部位置时，声像方位不会改变，这一点与实际情况不符。这也是仿真头立体声重放时，特别当用耳机重放普通声级差式立体声时，声像出现在头内的原因之一。为了避免头中定位效应，有必要对头部移动进行跟踪，同时更新双耳信号，使系统的输出能够随头部方位的变化而变化。这种信号处理技术在虚拟现实系统中是非常必要的。因为在虚拟现实系统中，声源的方位感和视觉看到的信息是密切关联的，必须相互一致。为了做到这一点，一种方法是测量出不同方向的双耳滤波器频率响应，这些滤波器函数称为头部相关函数（Head Related Transfer Functions，缩写为 HRTFs），如图 7.30 所示。这些滤波器函数也可以用于仿真头立体声信号处理系统中，将某个单声道声音定位于指定方向。由于在不同方向之间用插入法计算某个方向的 HRTFs 比较困难，因此往往需要测量出不同水平方位角和垂直方位角的 HRTFs，组成一个巨大的数据库存储起来，以便使用。当对仿真头录音信号进行处理时，HRTFs 数据必须根据头部实际方位进行修正，才能达到跟踪头部位置的目的。

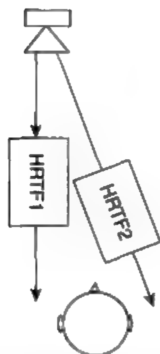


图 7.30　仿真头立体声的头部相关函数（HRTFs）

普通双声道立体声用耳机重放会产生一定的声场畸变，而仿真头立体声也不能用扬声器系统准确地重放。原因是在前一种情况下，声音重放时头部的衍射效应被去除了；而在后一种情况下，头部对声波的衍射效应作用了两次，一次是在录音的时候，另一次是在重放的时候。利用头部相关函数可以将信号从一种重放制式转换到另一种重放制式。

为了将普通立体声信号转换成适合耳机重放的仿真头立体声信号，需要将两路信号分别经过与扬声器和头部的相对位置有关的头部相关函数的滤波处理，如图 7.31 所示。请注意这里需要使用 4 个滤波器进行处理，因为左、右两只扬声器的声音不仅分别到达左、右耳，而且会互相交叉到达另一只耳朵。

为了将仿真头信号转换成适合普通双声道立体声系统重放的信号，需要将重放扬声器方向的头部相关函数从信号中去除，如图 7.32 所示。这样的信号处理要用到两个 HRTFs 的反转函数和两个模拟交叉信号传输特性的滤波器，才能在重放时将这些额外信号抵消掉。显然，这样的系统只能针对特定的扬声器摆位和听音者位置而设计，而且听音时头部不能移动，因此实用性不强。但是，这项技术已经在某些增强声音空间感的专利技术中得到了应用。

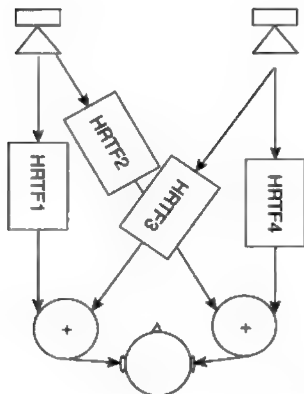


图 7.31　从立体声信号到仿真头信号的转换

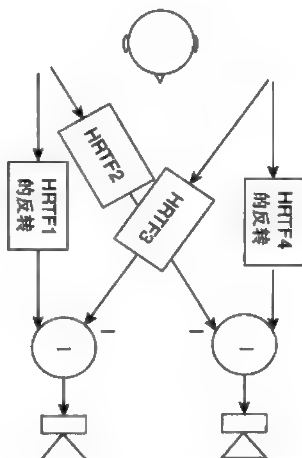


图 7.32　从仿真头信号到立体声信号的转换

7.8　响度处理

有时我们需要控制声音信号的电平，通常是修正声音信号的动态范围，使其更适合于特定的音乐内容。满足这类需求的两种主要信号处理设备称为压限器和扩展器（或噪声门）。这两种设备的功能都是首先检测信号的幅度，然后用信号幅度控制某种形式的增益控制装置。在数字信号处理中，一个简单的乘法运算就可以满足这个要求。

压限器和扩展器（或噪声门）的主要区别是：

（1）压限器：这类装置用来减小声音信号的动态范围，使较大的动态

范围变得较小。可以在整个信号动态范围内以相同的比率减小信号幅度，但更常见的是在设定的信号电平或阈值之上进行处理。一般控制参数是阈值电平和动态范围的压缩量，后者通常称为压缩比。1 ∶ 1 的压缩比意味着没有压缩；2 ∶ 1 的压缩比意味着有一定压缩；20 ∶ 1 的压缩比意味着极大的压缩。当压缩比更大时，输出信号电平几乎不跟随输入信号电平变化，这种信号处理通常称为限制。典型压缩特性曲线如图 7.33 所示。

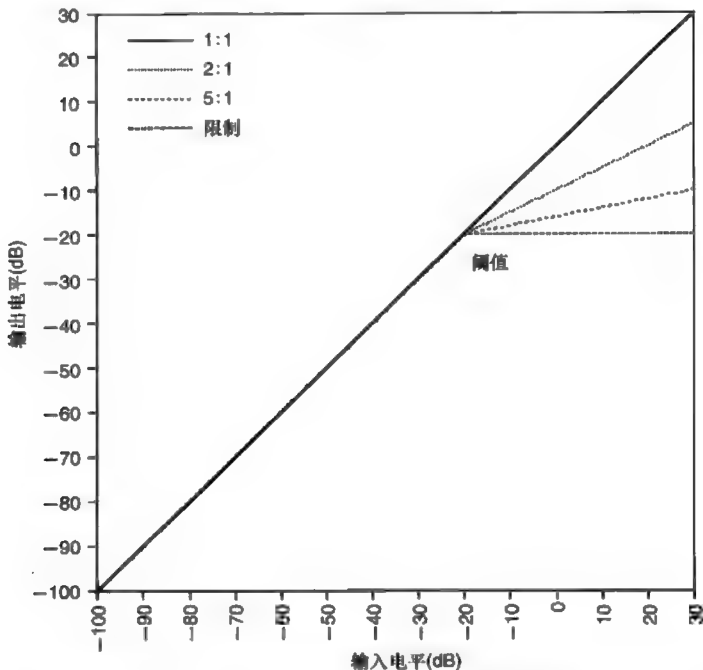


图 7.33　压限器输出输入特性曲线

压限器的作用之一是提高声音的响度，因为它可以增大信号的平均功率，而平均功率和响度感觉的关系最为密切。它的另一个主要作用是控制声音信号动态范围的峰值，防止信号产生失真。

对压限器而言，重要的是在使用时不能产生任何可察觉的人为声音，如音量的突然变化、重击声、信号本底噪声的起伏或信号失真，这需要对压限器的时间常数进行精心设置。通常使用者可以对压限器的启动速度即建立时间和信号幅度减小后压限器停止压缩的速度即恢复时间进行控制。一个可行的操作准则是根据信号的混响时间设置恢复时间，这样可以使信号改变幅度的时间得到较好的控制。

（2）扩展器或噪声门：这类装置主要用来增大声音信号的动态范围，使较小的动态范围变得较大。可以在整个信号动态范围内以相同的比率增大信号幅度，但更常见的是在设定的信号电平或阈值之下进行处理。一般控制参数是阈值电平和动态范围的扩展量，后者通常称为扩展比。1 ：1 的扩展比意味着没有扩展；1 ：2 的扩展比意味着有一定扩展；1 ：20 的扩展比意味着极大的扩展。极大的扩展使得当输入信号电平低于阈值时，输出信号几乎为零，这种处理称为门限。通常也有一种方法用于限制增益减小量，称为范围控制。典型扩展特性曲线如图 7.34 所示。

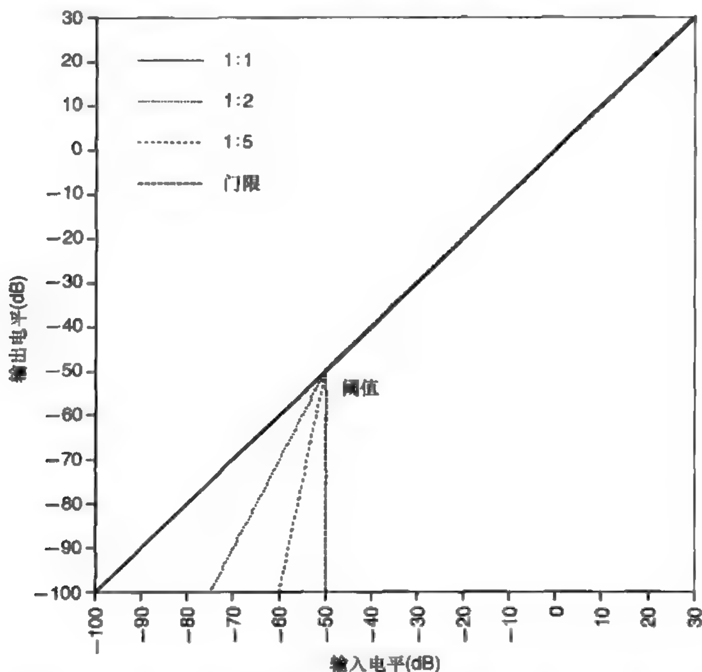


图 7.34 扩展器输出输入特性曲线

扩展器处理后会使声音听起来响度减小，因为信号的平均功率降低了。但是，扩展器的主要作用是降低所录制信号的噪声，因为当将阈值设定在噪声电平时，当信号电平低于阈值电平时，就会被自动减弱。这种声音信号处理会减小噪声对主观听音的影响，这是因为当声音信号较大时，噪声会被声音信号所掩蔽，而当声音信号较小或处于间隙阶段时，噪声会被自动减弱。

为了使扩展器正常工作，也需要设定合适的建立时间和恢复时间。

建立时间决定当信号电平大于阈值时增益提高的速度，应该足够短。对于某些声音如打击乐，建立时间越短越好。但是，当用较短的建立时间处理起振阶段较长的声音时，会生产可察觉的咔嗒声。因此，一般这个时间常数要根据具体情况进行调节。

当声音的起振时间较短时，建立时间应该设置在一个较小值；当声音的起振时间较长时，这个时间常数就应该较大一些。恢复时间是指当信号电平低于阈值时增益减小的速度。像压限器一样，恢复时间一般设置在与信号混响时间相当的值。

这两种信号处理设备都可以选择这样的一种工作状态，即阈值不是一个精确的"转折点"，而是在一定电平范围的逐渐过渡。这种逐渐过渡的信号处理方式可以避免输出信号产生可察觉的人为声。

虽然响度处理设备经常用于控制声音的动态范围和噪声，它们也可以创造性地用于修正声音信号的包络，从而产生一个新的声音。这种技术也可以用于产生立体声效果，如自动声像调整，使声像在立体声重放舞台动态移动。

7.9 总结

本章介绍了用电子设备处理声音信号的一些方法及其背后的心理声学原理。还有许多其他声音信号处理方法是以人耳听觉特性为基础的，如降噪、音频压缩编码等。事实上，心理声学是所有声音信号处理技术的基础，因为就像其他科学一样，物理声学解释正在发生的现象，而心理声学解释听觉怎样听到声音，因此最终决定了我们解决问题的具体方法。声学和心理声学相结合，共同成为人类努力创造良好的音乐体验和音乐艺术表现的基础。

参考文献

Bartlett, B. (1970). A scientific explanation of phasing (flanging). Journal of the Audio Engineering Society, 18, (6), 674.

Bohn, D.A. (1986). Constant-Q graphic equalizers. Journal of the Audio Engineering Society, 34, (9), 611.

Bohn, D.A. (1992). Accelerated slope tone control equalizers. Journal of the Audio Engineering Society, 40, (12), 1018.

Bristow-Johnson, R. (1995). A detailed analysis of a time-domain formant-corrected pitch-shifting algorithm. Journal of the Audio Engineering Society, 43, (5), 340.

Dolby, R.M. (1967). An audio noise reduction system. Journal of the Audio Engineering Society, 15, (4), 383.

Dolby, R. (1983). A 20 dB audio noise reduction system for consumer applications. Journal of the Audio Engineering Society, 31, (3), 98.

Greiner, R.A. and Sohoessow, M. (1983). Design aspects of graphic equalizers. Journal of the Audio Engineering Society, 31, (6), 394.

Hartmann, W.M. (1978). Flanging and phasers. Journal of the Audio Engineering Society, 26, (6), 439.

Henriquez, J.A., Riemer, T.E., and Trahan, R.E. Jr (1990). A phase-linear audio equalizer: design and implementation. Journal of the Audio Engineering Society, 38, (9), 653.

Kahrs, M. and Brandenburg, K. (eds) (1998). Applications of Digital Signal Processing to Audio and Acoustics. The Kluwer International Series in Engineering and Computer Science, Vol. 437. Boston and London: Kluwer Academic.

Kraght, P.H. (1992). A linear-phase digital equalizer with cubic-spline frequency response. Journal of the Audio Engineering Society, 40, (5), 403.

Lee, F.F. (1972). Time compression and expansion of speech by the sampling method. Journal of the Audio Engineering Society, 20, (9), 738.

Moorer, J.A. (1978). The use of the phase vocoder in computer music applications. Journal of the Audio Engineering Society, 26, (1), 42.

Oliveira, A.J. (1989). A feedforward side-chain limiter/compressor/deesser with improved flexibility (ER). Journal of the Audio Engineering Society, 37, (4), 226.

Orfanidis, S.J. (1997). Digital parametric equalizer design with prescribed Nyquist-frequency gain (P). Journal of the Audio Engineering Society, 45, (6), 444.

Rumsey, F. (2001). Spatial Audio, Oxford: Focal Press.

Schroeder, M.R. (1962). Natural sounding artificial reverberation. Journal of the Audio Engineering Society, 10, (3), 219.

Stikvoort, E.F. (1986). Digital dynamic range compressor for audio. Journal of the Audio Engineering Society, 34, (1), 3.

Z.lzer, U. (1997). Digital Audio Signal Processing, Chichester: Wiley.

附录 1　求解 *ERB* 等式

为了求出某已知临界频带带宽或 *ERB* 的中心频率，可将联系 *ERB* 和中心频率 f_c 的式子（式（2.6））做如下变形。

式（2.6）为：

$$ERB = \{24.7 \times [(4.37 \times f_c)+1]\}\ \text{Hz} \tag{A1.1}$$

其中，f_c 为滤波器中心频率（kHz），*ERB* 为等效矩形滤波器带宽（Hz）。等式有效频率范围是 100Hz < f_c < 10 000Hz。

将上式逐步变形，可推导出由 *ERB* 计算 f_c 的公式，过程如下：

$$\frac{ERB}{24.7} = (4.37 \times f_c)+1$$

因此：

$$\frac{ERB}{24.7} - 1 = (4.37 \times f_c)$$

所以：

$$f_c = \left[\frac{\left(\frac{ERB}{24.7}\right)-1}{4.37}\right]\text{kHz} \tag{A1.2}$$

附录 2　频率比与音分之间的换算

音分定义为平均律半音音程的百分之一。由于 1 个倍频程里有 12 个半音，因此，一个音分等于一个倍频程的 $\frac{1}{1\,200}$，即：

$$\sqrt[1\,200]{2} \text{ 或 } 2^{\left[\frac{1}{1\,200}\right]}$$

因此，F_1 与 F_2 之间的频率比可以由音分值计算如下：

$$\frac{F_1}{F_2}=2^{\left[\frac{c}{1\,200}\right]}$$

将上式变形后，可得到由频率比计算音分值的公式：

$$\log_2\left[\frac{F_1}{F_2}\right]=\left[\frac{c}{1\,200}\right]$$

即

$$c=1\,200\log_2\left[\frac{F_1}{F_2}\right] \qquad (\text{A2.1})$$

为了计算方便，可以将以 2 为底的对数变换为以 10 为底的对数。设：

$$\lg 2^{(x)}=y \qquad (\text{A2.2})$$

则

$$x=2^{y}$$

两边取 10 为底的对数得：

$$\lg 10^{(x)}=\lg 10^{(2^y)}=y\lg 10^{2}$$

将上式代入式（A2.2）中的 y 得：

$$\log_2[x]=\left[\frac{\lg[x]}{\lg 2}\right] \tag{A2.3}$$

再将式（A2.3）代入式（A2.1）得：

$$c=1\,200\left\{\frac{\lg\left[\frac{F_1}{F_2}\right]}{\lg 2}\right\}=\left[\frac{1\,200}{\lg 2}\right]\lg\left[\frac{F_1}{F_2}\right]$$

计算常数项后，得到从频率比计算音分值的公式为：

$$c=3\,986.313\,7\lg\left[\frac{F_1}{F_2}\right] \tag{A2.4}$$

因此，半音数（s）可用下式计算：

$$s=\left[\frac{c}{100}\right]=39.863\,137\lg\left[\frac{F_1}{F_2}\right] \tag{A2.5}$$

由式（A2.4）可得从音分值计算频率比的公式：

$$\frac{F_1}{F_2}=10^{\left[\frac{c}{3\,986.313\,7}\right]} \tag{A2.6}$$

附录 3　混响时间公式的推导

声场衰减所需的时间显然不仅与房间表面的吸声性能有关，而且与房间内声波相邻两次撞击墙面的时间间隔有关。利用这些因素可以推导出房间混响时间计算公式。首先要计算两次相邻反射的平均时间间隔，这可以从平均自由程（Mean free path）计算出来。平均自由程是假设声波到达各个方向和各个位置机率相同的情况下，声波在两个界面之间传播的平均路程。对于矩形房间，平均自由程由下式给出：

$$MFP=\frac{4V}{S} \tag{A3.1}$$

其中，MFP 为平均自由程（m），V 为房间体积（m^3），S 为房间总表面积（m^2）。

将式（A3.1）除以声速，可得两次反射的时间间隔为：

$$\tau=\frac{4V}{Sc} \tag{A3.2}$$

其中，τ 为两次反射的时间间隔（s），c 为声速（m/s）。

式（A3.2）给出了两次反射的时间间隔。在每次反射时，都有 α 部分的声能被壁面吸收，α 为前面介绍的平均吸声系数。如果 α 部分的能量被壁面吸收，那么就有（$1-\alpha$）部分的能量被反射，并作用于下一个壁面。而且每一次反射总有 α 部分声能被进一步吸收，因此，反射回来的声能为：

$$
\begin{aligned}
&\text{1 次反射后} && E_1=E_0(1-a)\\
&\text{2 次反射后} && E_2=E_0(1-a)^2\\
&\text{3 次反射后} && E_3=E_0(1-a)^3\\
& && \vdots\\
&n\text{ 次反射后} && E_n=E_0(1-a)^n
\end{aligned}
\tag{A3.3}
$$

由于 a 为小于 1 的系数，（1-a）也小于 1，因此式（A3.3）表明声能按指数规律衰减。为了求出声能衰减到指定比例所需的时间，就要先求出在指定时间间隔内声波发生反射的次数，而这可以通过将这个时间间隔除以两次反射的间隔得到。利用式（A3.2）得：

$$n=\frac{t}{\left(\frac{4V}{Sc}\right)}=t\left(\frac{Sc}{4V}\right) \tag{A3.4}$$

其中，t 为时间间隔（s）。

将式（A3.4）代入式（A3.3）后，得到经过一定时间 t 后所剩余的声能为：

$$E_t=E_0(1-a)^{t\left(\frac{Sc}{4V}\right)} \tag{A3.5}$$

因此，经过一定时间 t 后，声能的衰减比例为：

$$\frac{E_t}{E_0}=(1-a)^{t\left(\frac{Sc}{4V}\right)} \tag{A3.6}$$

为了计算达到指定衰减比例所需的时间，对式（A3.6）两边取以（1-a）为底的对数得：

$$\log_{(1-a)}\left(\frac{E_t}{E_0}\right)=t\left(\frac{Sc}{4V}\right)$$

因此，达到一定衰减比例所需时间为：

$$t=\left(\frac{4V}{Sc}\right)\log_{(1-a)}\left(\frac{E_t}{E_0}\right) \tag{A3.7}$$

将以（1-a）为底的对数转换为自然对数得：

$$t=\left(\frac{4V}{Sc}\right)\frac{\ln\left(\frac{E_t}{E_0}\right)}{\ln(1-a)} \tag{A3.8}$$

式（A3.8）给出了声能衰减比例与所需时间的关系，可用于计算所需的时间。我们可以选择无限多种比例进行计算，但是最常用的是声能衰

减 60dB 的声能比，即 10^{-6}。将此值代入式（A3.8），得到声能衰减 60dB 所需的时间，即混响时间 T_{60} 为：

$$T_{60}=\left(\frac{4V}{Sc}\right)\frac{\ln\left(10^{-6}\right)}{\ln(1-\alpha)}=\left(\frac{V}{S\ln(1-\alpha)}\right)\frac{4\times(-13.82)}{344}=\frac{-0.161V}{S\ln(1-\alpha)} \tag{A3.9}$$

其中，T_{60} 为 60dB 混响时间（s）。

因此，混响时间计算公式为：

$$T_{60}=\frac{-0.161V}{S\ln(1-\alpha)} \tag{A3.10}$$

式（A3.10）称为艾润混响公式，式中负号与自然对数计算结果的负号相抵消，得到一个正的混响时间值。在实际应用中，也可能会遇到计算其他衰减比例的混响时间，其计算公式与式（A3.10）的唯一不同是常数部分。上述公式的推导是基于统计学，因此式（A3.10）的应用必须具备以下前提条件：

（1）声波以相同的机率从各个方向到达墙面的各个位置，即声场为扩散声场。这个条件是为了利用房间平均吸声系数的概念。请注意从主观听音上看，这也是所期望达到的目标，人们更愿意在扩散声场中聆听或演奏音乐。

（2）平均自由程概念成立。这也是利用平均吸声系数的条件。此外，它也意味着房间的形状不能太极端，即这种分析方法不适用于类似长隧道的房间形状。然而，对于大多数实际房间，平均自由程的计算公式是适用的。

附录4　不同频率和表面材料混响时间公式的推导

在实际房间中，我们不仅要考虑到存在不同吸声材料的可能性，而且还要考虑到吸声系数随频率的变化特性。由于不同吸声材料的面积不同，因此要找到一种能够反映不同吸声材料相对吸声大小的方法就变得比较复杂。例如，吸声系数较小而面积较大的吸声面的影响可能大于吸声系数较大而面积很小的吸声面。在赛宾公式中，解决这个问题的方法比较简单，只要将吸声系数乘以相应的吸声面积，然后将房间内所有表面的这个乘积结果相加即可。Sabine 将这个量称为“等效开窗面积”。他假设并通过实验验证了开窗的吸声系数等于 1。因此，较为简便的方法是将上述计算得到的“等效开窗面积”代替赛宾公式中用平均吸声系数计算的等效开窗面积，得到房间里存在不同吸声材料、吸声系数随频率变化的混响时间修正公式：

$$T_{60(a<0.3)}=\frac{0.161V}{\sum_{S_i} S_i a_i(f)} \qquad (A4.1)$$

其中，$a_i(f)$ 为给定材料的吸声系数，S_i 为其面积。

在艾润混响公式中，由于没有直接用到等效开窗面积，所以情况会稍微复杂一些。通常有两种方法，一种是先计算出等效开窗面积，再除

以总面积，得到加权平均吸声系数。因此，加权平均吸声系数的计算公式为：

$$a(f)=\frac{\sum_{S_i} S_i a_i(f)}{S}$$

将上式代入艾润混响公式，得到不同吸声材料的修正公式：

$$T_{60}=\frac{-0.161V}{S\ln\left(1-\frac{\sum_{S_i} S_i a_i(f)}{S}\right)} \tag{A4.2}$$

式（A4.2）可用于计算不同吸声材料、吸声系数随频率变化的房间混响时间。但是另一种方法更符合艾润混响公式的推导理论。这种方法通过分析声波经过不同吸声系数表面连续反射后的声能进行推导。在这种情况下，声波每次反射回来的声能所占入射声能的比例各不相同：

1 次反射后　$E_1=E_0(1-a_1)$

2 次反射后　$E_2=E_0(1-a_1)(1-a_2)$

3 次反射后　$E_3=E_0(1-a_1)(1-a_2)(1-a_3)$

⋮　⋮

n 次反射后　$E_n=E_0(1-a_1)(1-a_2)(1-a_3)\cdots(1-a_n)$

因此，可以找到一个平均吸声系数 a，它是不同反射系数（$1-a_i$）的几何平均值。例如，假设声波经过两次反射后声能等于经过两次相同反射面反射后的声能，则平均反射系数为：

$$1-a=\sqrt{(1-a_1)(1-a_2)}$$

三次不同反射的平均反射系数为：

$$1-a=\sqrt[3]{(1-a_1)(1-a_2)(1-a_3)}$$

n 次不同反射的平均反射系数为：

$$1-a=\sqrt[n]{(1-a_1)(1-a_2)(1-a_3)\cdots(1-a_n)}$$

虽然房间内不同表面的数目是一定的，但是它们却有不同的面积，因此不仅要对不同的吸声表面数目取平均，而且还要对不同的面积进行加权。由于各项乘积的对数等于各项对数之和，面积加权后的几何平均值可以转换为各项吸声之和。因此，当墙面吸声系数变化或者吸声系数随频率变化时，艾润混响时间计算公式修正为：

$$T_{60} = \frac{-0.161V}{\sum_{S_i} S_i \ln\left(1-\alpha_i(f)\right)} \quad (A4.3)$$

式（A4.3）也称为 Millington-Sette 公式。虽然式（A4.3）适用于任何吸声系数值，但它还是比赛宾公式更为复杂一些。前面提到过，当吸声系数小于 0.3 时，则两个公式的计算结果非常接近，可以用较简单的公式进行计算。因此，在许多情况下，式（A4.1）更为常用。

附录 5　CD 听音项目一览表

本书所附 CD 包含一系列与本书内容相关的听音资料，以便读者更好地理解和掌握书中阐述的内容。此 CD 仅限于购买此书者学习使用，版权属于 David M Howard。任何使用该 CD 从事商业活动的行为必须征得版权所有者的同意。

音乐声学和心理声学 CD 的版权所有者
David M Howard，2005 年于约克郡

下面所列的 CD 目录给出了各条听音材料所支持的章节，并对每项听音内容做出书面说明。其中有些听音项目与心理声学内容相关，会指出需要用耳机听音，并且重放音量要调整到较舒适的大小。有些听音项目是在消声室录制的一些声学乐器演奏的音乐，主要作为源材料，适合听音者用自己的听觉进行分析。例如，听音者可以根据乐音的波形和频谱不同进行主观听音比较，乐音的波形和频谱可参看书中图 4.11、图 4.17、图 4.22、图 4.29 和图 5.1 到图 5.5。一些这方面的免费应用软件可以利用因特网搜索引擎从网上下载。

CD 的录音是在英国约克大学电子系的消声室进行的，母带录制由一支 Sennheiser MKH20 全指向话筒、一台 RME 四声道话筒放大器和一台

Edirol R4 硬盘录音机（采样频率 44kHz、分辨率 24bit）完成。值得注意的是，在录音过程中没有刻意去除乐音中的一些额外声音，如呼吸声、乐器按键声和乐谱翻动声等，这些噪声被保留在源材料中，以便读者在一些实际操作练习中使用，比如学习如何去除这些额外声音等。

表中 No. 表示听音项目的编号，章节号表示该听音项目与书中这个部分内容有关。如果该听音项目与书中某个图有关，那么这些信息会在项目内容中体现出来。

《音乐声学和心理声学》CD 听音项目表

No.	内　　容	章节
1	临界频带：两个纯音同时播放，一个频率（F1）固定在 440Hz，另一个频率（F2）从 440Hz 逐渐增大到 660Hz，再从 660Hz 逐渐减小到 440Hz。感受书中第 2.2 节和图 2.6 所描述的听音感觉的变化。	2.2
2	年龄老化引起听力损失：No.73 中众赞歌前奏曲的若干小节经过滤波器处理后产生年龄老化引起听感变化的效果。滤波器处理以图 2.11 为依据，年龄依次为 20、60 和 80（岁）。特别注意为众赞歌定音的曲调（详见 No.73）的变化。	2.3
3	响度加倍：为使响度增大 1 倍，声音的声压级大约需增大 10dB，参看例 2.4。节目为消声室录音，相对声压级设置在 0dB、+3dB、+6dB 和 +10dB，以测试声强加倍（+3dB）、声压加倍（+6dB）和响度加倍（+10dB）产生的响度变化感觉。	2.4
4	音高展示：音高展示包括以下内容。A：基频 200Hz 及其最低的 5 个谐波；B：和 A 基本相同，但不包含基频（缺失基频的情况）；C：由 1 800Hz、2 000Hz 和 2 200Hz 组成的声音，听觉往往能够感觉到 200Hz 的音高；D：由 1 840Hz、2 040Hz 和 2 240Hz 组成的声音，听觉能模糊地感觉到 207Hz 的音高。对（C）和（D）听音项目的解释，可参看书中 3.2.1 节，参考文献（Schouten 1940 和 Moore 1982）见第 3 章后面所列。	3.2.1 3.2.2 3.2.3
5	质数频率分量：由质数频率分量 557Hz、1 381Hz、1 663Hz、1 993Hz 和 2 371Hz 合成的声音，因此它们不是某个共同基频的谐波。	3.2.2 5.3.1
6	重复音高：用噪声性声音与其延迟声叠加后演奏琶音和滑音，详见第 3.2.6 节的最后一个段落。	3.2.6
7	音律：一个和弦系列（分别是 C 大调、B 大调和 F 大调）按以下音律顺序展示：A、五度相生律（以 C 键为根音）；B、纯律（以 C 键为根音）；C、十二平均律。	3.4
8	1/2 尺寸小提琴琶音：由 1/2 尺寸小提琴演奏的 G 大调琶音（从 G3 到 G5，再回到 G3）。	4.2
9	1/2 尺寸小提琴音乐片断：由 1/2 尺寸小提琴演奏的音乐片断。	4.2
10	小提琴琶音：小提琴演奏的 G 大调琶音（从 G3 到 G7，再回到 G3）。	4.2
11	长笛琶音：长笛演奏的 G 大调琶音（从 D4 到 B6，再回到 D4）。	4.3.3
12	长笛音乐：长笛演奏的音乐片断。	4.3.3
13	短笛琶音：短笛演奏的 G 大调琶音（从 D5 到 B7，再回到 D5）。	4.3.3
14	短笛音乐：短笛演奏的音乐片断。	4.3.3
15	低音竖笛琶音：低音竖笛演奏的 G 大调琶音加极限音（从 F2、G2 到 G4 再回到 F2）。	4.3.3
16	低音竖笛音乐：低音竖笛演奏的音乐小片断。	4.3.3
17	中音竖笛琶音：中音竖笛演奏的 G 大调琶音加极限音（从 C3、D3 到 B4、C5 再回到 C3）。	4.3.3
18	中音竖笛音乐：中音竖笛演奏的音乐片断。	4.3.3
19	高音（treble）竖笛琶音：高音竖笛演奏的 G 大调琶音加极限音高音（从 F3、G3 到 G5，再回到 F3）。	4.3.3

续表

No.	内　　容	章节
20	高音（treble）竖笛音乐：高音竖笛演奏的音乐片断。	4.3.3
21	和音竖笛琶音：和音竖笛演奏的 G 大调琶音加极限音（从 C4、D4 到 B5、C6，再回到 C4）。	4.3.3
22	和音竖笛音乐：和音竖笛演奏的音乐片断。	4.3.3
23	高音（sopranino）竖笛琶音：高音竖笛演奏的 G 大调琶音加极限音（从 F4、G4 到 D6，再回到 F4）。	4.3.3
24	高音（sopranino）竖笛音乐：高音竖笛演奏的音乐片断。	4.3.3
25	斯旺尼哨：斯旺尼哨的快速和慢速滑奏。	4.3.3
26	巴松管（低音双簧管）琶音：巴松演奏的 G 大调琶音加极限音（从 B♭1、B1 到 B4，再回到 B♭1）。	4.3.6
27	巴松音乐：巴松演奏的音乐片断。	4.3.6
28	双簧管琶音：双簧管演奏的 G 大调琶音（从 B3 到 D6，再回到 B3）。	4.3.6
29	双簧管音乐：双簧管演奏的音乐片断。	4.3.6
30	风笛和弦：持续和低沉的风笛和弦。	4.3.6
31	风笛音乐：风笛演奏的音乐片断。	4.3.6
32	低音单簧管琶音：低音单簧管演奏的 G 大调琶音（从 D2 到 G4，再回到 D2）。	4.3.6
33	低音单簧管音乐：低音单簧管演奏的音乐片断。	4.3.6
34	单簧管琶音：单簧管演奏的 G 大调琶音（从 D3 到 G5，再回到 D3）。	4.3.6
35	单簧管音乐：单簧管演奏的音乐片断。	4.3.6
36	土巴（大号）琶音：大号演奏的 G 大调琶音（从 G1 到 G3，再回到 G1）。	4.3.7
37	大号音乐：大号演奏的音乐片断。	4.3.7
38	鸟鸣声：布谷鸟（3 次）	4.3.3
39	鸟鸣声：野鸭（8 次）	4.3.6
40	鸟鸣声：鹅（7 次）	4.3.6
41	鸟鸣声：夜莺（5 次）	4.3.3
42	鸟鸣声：鹌鹑（8 次）	4.3.3
43	成人女声：朗读某段落（左右声道都由传声器录制）	4.5
44	成人女声：朗读某段落（右声道：传声器；左声道：喉电描记器 *（electrolaryngograph，缩写为 Lx））	4.5
45	成人女声琶音：G 大调琶音（从 B3 到 B5，再回到 B3。两个声道：传声器）	4.5
46	成人女声琶音：G 大调琶音（从 B3 到 B5，再回到 B3。右声道：传声器；左声道：喉电描记器 *）	4.5
47	成人女声歌唱：歌唱片断（两个声道：传声器）	4.5
48	成人女声歌唱：歌唱片断（右声道：传声器；左声道：喉电描记器 *）	4.5
49	成人男声：朗读某段落（左右声道都由传声器录制）	4.5
50	成人男声：朗读某段落（右声道：传声器；左声道：喉电描记器 *）	4.5
51	成人男声琶音：G 大调琶音（从 G2 到 B4，再回到 G2。两个声道：传声器）	4.5
52	成人男声琶音：G 大调琶音（从 G2 到 B4，再回到 G2。右声道：传声器；左声道：喉电描记器 *）	4.5
53	成人男声歌唱：四声部多声道男声四重唱片断（TTBB）（传声器录音的立体声重放）	4.5
54	成人男声歌唱：四声部多声道男声四重唱片断（TTBB）（喉电描记器输出的立体声重放）	4.5
55	成人男声歌唱：四声部多声道男声四重唱片断（TTBB）（第一高音由传声器录制，其他三个声部由喉电描记器记录）	4.5
56	女童（9 岁）歌声：赞美诗片断（两个声道：传声器）	4.5
57	女童（9 岁）歌声：赞美诗片断（右声道：传声器；左声道：喉电描记器 *）	4.5

续表

No.	内　　容	章节
58	男童（12 岁）语声：朗读某段落（左右声道都由传声器录制）	4.5
59	男童（12 岁）语声：朗读某段落（右声道：传声器；左声道：喉电描记器*）	4.5
60	男童（12 岁）歌声：赞美诗片断（两个声道：传声器）	4.5
61	男童（12 岁）歌声：赞美诗片断（右声道：传声器；左声道：喉电描记器*）	4.5
62	音色展示：4 个音符分别由小提琴、长笛、巴松管和双簧管演奏，各音音高为 G4，起始阶段振幅上升、结束阶段振幅下降，稳态阶段调整后使音量相当。因为各音的音高、响度和持续时间相同，因此，各音在听感上的差异就是音色差（参看第 5.1 节给出的音色定义）。	5.1
63	Helmholtz 音色规律：一系列声音用来展示 Helmholtz 关于音色的 4 条"规律"（见 5.3.2 节）。A：200Hz 的正弦波（规律 1——纯音音色）。B：包含 200Hz、400Hz、600Hz、800Hz 和 1 000Hz 的谐波（定律 2——乐音）。C：包含 200Hz、600Hz、1 000Hz、1 400Hz、1 800Hz、2 200Hz、2 400Hz、2 600Hz、3 000Hz、3 400Hz 和 3 800Hz 的谐波（规律 3——不均匀分布的谐波）。D：包含基频为 200Hz 的 1 ~ 20 的谐波（规律 4——丰富的第 6 或第 7 次以上的高次谐波）。	5.3.2
64	加强所有谐波的管风琴音栓：保持音符 G4（392Hz），依次拉起加强前 9 个谐波成分（$1f_0$、$2f_0$、$3f_0$、$4f_0$、$5f_0$、$6f_0$、$7f_0$、$8f_0$、$9f_0$）的音栓。这个乐器上的音栓有：chimney flute 8'、open flute 4'、nazard $2\frac{2}{3}'$、block flute 2'、tierce $1\frac{3}{5}'$、larigot $1\frac{1}{3}'$、septième $1\frac{1}{7}'$ 和 none $\frac{8}{9}'$。通过这个演示可以听到每个音栓对音色产生的影响，叫做"分析"听音。然后音符被重放 3 次，并且演奏一组琶音，使音符作为一个整体被听到，叫做"整体"听音。	5.4
65	加强奇次谐波的管风琴音栓：保持音符 G4（392Hz），依次拉起加强前 5 个奇次谐波（$1f_0$、$3f_0$、$5f_0$、$7f_0$、$9f_0$）的音栓。这个乐器上的音栓有：chimney flute 8'、nazard $2\frac{2}{3}'$、tierce $1\frac{3}{5}'$、septième $1\frac{1}{7}'$ 和 none $\frac{8}{9}'$。通过这个演示可以听到每个音栓对音色产生的影响，叫做"分析"听音。然后音符被重放 3 次，并且演奏一段旋律，使音符音色作为一个整体被听到，叫做"整体"听音。注意最后听到的声音具有单簧管音色，因为声音包含奇次谐波系列。	5.4
66	听觉的相近频率编组效应：图 5.13 所示乐谱被演奏后，可听到编组效应。该乐谱选自 JS Bach 的 E 大调组曲 III 中前奏曲的片断。	5.5.3
67	编组和相近频率编组效应：柴可夫斯基第六交响乐最后一个乐章的弦乐片断（乐谱见图 5.14）演奏后，可以听到以下编组效应。A：整个乐队。B：第一小提琴和中提琴。C：第二小提琴和大提琴。	5.5.3
68	"Shepherd 音"幻觉：基于图 5.18 所示频谱的"Shepherd 音"幻觉演示。	5.5.4
69	"Shepherd 音"构成元素：演示上一条音轨中各个声音成分对"Shepherd 音"的循环听音幻觉的作用。图 5.18 显示各个声音成分是怎样被控制的。	5.5.4
70	连续音阶音高幻觉 1：演示图 5.19 所示的 J.S.Bach 的管风琴曲《G 小调幻想曲》（BWV 542）片断，在键盘部分不使用簧管音栓。看着乐谱，通过判断音符跳跃到哪个八度，判断向上的七度跳跃的音高感觉是否模糊。	5.5.4
71	连续音阶音高幻觉 2：演示图 5.19 所示的 J.S.Bach 的管风琴曲《G 小调幻想曲》（BWV 542）片断，在键盘部分使用簧管音栓。看着乐谱，通过判断音符跳跃到哪个八度，判断向上的七度跳跃的音高感觉是否模糊。	5.5.4
72	伪造的连续音阶音高：演示图 5.19 所示的 J.S.Bach 的管风琴曲《G 小调幻想曲》（BWV 542）片断，键盘部分用记号 1 演奏（不使用簧管音栓），但是在演奏过程中改变音栓，在踏板的连续下降部分用簧管音栓。具体作法是：第一个下降八度用 schalmei 4'、第二个下降八度用 trumpet 8'、第三个下降八度用 ophecleide 16'、最后一个下降七度用 contra bombarde 32'。需要注意的是，这些簧管具有不同的音色和强度，因此音栓的变换可以清楚地听出来。尽管如此，长的下降音阶还是被明确表示出来。	5.5.4

续表

No.	内　容	章节
73	音高幻觉——普通演奏的众赞歌：根据5.5.4节最后一段引用的Roederer的建议，这个演示采用J.S.Bach的管风琴曲集的众赞歌前奏曲，该曲目作为众赞歌第41号记录在Novello版本（Ivor Atkins 编辑，Walter Emery修订）中。Roederer所提议的众赞歌第40号，在Novello版本中没有独奏的众赞歌旋律，因此不适合做这个演示。在这个演示中，众赞歌片断的旋律用加强前9个谐波的音栓（chimney flute 8'、open flute 4'、nazard 2⅔'、block flute 2'、tierce 1⅗'、larigot 1⅓'、septi è me 1⅐'、octavin 1'和none ⁸⁄₉'）演奏（在一般演出中可能不会采用，但在这里可以加强演示效果）。	5.5.4&3.2.2
74	音高幻觉——管风琴众赞歌：在这个演示中仍然采用第73条的曲目，开始时使用加强前9个谐波的同样音栓，但是按以下顺序依次按下音栓：chimney flute 8'、open flute 4'、nazard 2⅔'、block flute 2'、tierce 1⅗'、larigot 1⅓'、septi è me 1⅐'和octavin 1'，直到只剩下none ⁸⁄₉'（加强第九次谐波），然后除去伴奏部分，随后再加入伴奏。虽然加强低次谐波的音栓被去除，曲调仍然能够听到，甚至当只剩下一个谐波（第9次）时也能听到曲调，这是转位（基频不是整数个八度变化）。当去除伴奏后，旋律有些模糊。	5.5.4&3.2.2
75	短脉冲响应：这是Dr.Damian Murphy提供的小型教堂脉冲响应，可用于卷积产生短混响RT_{60}效果。	6.1.4
76	中脉冲响应：这是Dr.Damian Murphy提供的中型教堂脉冲响应，可用于卷积产生中混响RT_{60}效果。	6.1.4
77	长脉冲响应：这是Dr.Damian Murphy提供的大空间脉冲响应，可用于卷积产生长混响RT_{60}效果。	6.1.4
78	单声道、立体声和环绕声：这项演示用耳机听音。通过从单声道到立体声再到环绕声，听音者感受声像如何定位于头外并发生变化。作者感谢约克大学电子系音乐技术研究小组成员提供了本项录音，它原本是为2001年7月题为"被声音环绕"的皇家协会夏季科学博览会而录制的。	7.7.4

* 喉电描记器（electolaryngograph or electroglottograph）是通过测量外置于喉部的两极之间的电阻抗来测量声带振动的一种装置。通过重放由喉电描记器（简称Lx）测得的波形，可得到发出语声和歌声的声源的相关信息。请注意Lx波形包含一些高频人工合成声，这些高频合成声通过重放可以尤其明显地听到，而通过低通滤波器可以将这些高频合成声去除（这个练习留给感兴趣的读者自己去做）。同时，由于滤波器相位特性的影响，任何滤波处理都可能使信号波形发生细微的变化。因此，Lx的输出信号是以原始的方式记录下来，以便需要时对波形做进一步的分析处理。通过测量Lx波形的基波周期（T_0），并通过公式$f_0=1/T_0$计算出基频，这是测量语声或歌声基频的精确和有效可行的实验方法。更多详细内容请参看第4章后面所列参考文献Baken（1987）、Howard（1995、1999）和Howard等人（1990）。

后　　记

经过一年的努力，终于完成了本书的翻译工作。在此首先感谢中国传媒大学在校研究生陈苇婧同学协助完成了第 1 章至第 3 章的第一稿翻译，她为此付出了大量的时间和精力。还要感谢本校音乐系的刘聪明老师，他帮助解决了翻译过程中遇到的一些音乐方面的问题。还要感谢中国传媒大学录音系领导和同事们对本书翻译工作的支持和帮助，以及富有建设性的讨论。以上这一切都为本书最终完成提供了必要的支持，在此表示衷心的感谢！

虽然尽最大努力完成了本书的翻译工作，但还不能做到完全满意，例如对书中的 5.4 节、7.5 节的翻译感到不够到位。因此希望广大读者包括同行、专家们对书中存在的不足之处提出宝贵意见和建议！

陈小平

2009 年 7 月于北京

[illegible]2 [illegible]
[illegible]3 [illegible]
[illegible]4 [illegible]
[illegible]5 CD[illegible]
[illegible]